# MODELING DIRECTIONAL KNOWLEDGE AND REASONING IN ENVIRONMENTAL SPACE: TESTING QUALITATIVE METRICS

**Daniel R. Montello and Andrew U. Frank**

**Abstract:** Researchers from a variety of disciplines have proposed models of human spatial knowledge and reasoning in order to explain spatial behavior in environmental spaces, such as buildings, neighborhoods, and cities. A common component of these models is a set of hypotheses about the geometry of spatial knowledge, particularly with respect to the roles of topological and metric knowledge. Recently, mathematicians and computer scientists interested in formally modeling everyday intelligent spatial behavior have developed models incorporating "qualitative" spatial reasoning ("naive" spatial reasoning). One branch of this effort has been the development of so-called "qualitative metric" models to solve problems such as wayfinding. A qualitative metric employs more sophisticated geometry than just topology but at a relatively imprecise or coarse-grained level. Such models essentially reason with a small finite number of quantitative categories for direction and/or distance. In this chapter, we evaluate the abilities of qualitative metric models to account for human knowledge of directions by comparing simulations derived from qualitative metrics to empirical data and theorizing derived from human-subjects testing.

## Introduction

Researchers have proposed models of spatial knowledge and reasoning in order to explain human spatial behavior in environmental spaces (the relatively large-scale spaces of buildings, neighborhoods, and cities). Although varying in comprehensiveness, these models have typically included ideas about processes of knowledge acquisition, the form of stored knowledge, and its retrieval and manipulation in working-memory. Such models have been provided by researchers from a wide variety of disciplines: geographers (Couclelis et al., 1987), psychologists (Piaget and Inhelder, 1948/1967; Siegel and White, 1975), and computer scientists (Kuipers and Levitt, 1988; McDermott and Davis, 1984), among others.

A common aspect of these models is a set of hypotheses about the geometric sophistication of spatial knowledge acquired from direct locomotor experience in the environment (cf. Golledge and Hubert, 1982; Kuipers and Levitt, 1988; Landau et al.,



*J. Portugali (ed.), The Construction of Cognitive Maps*, 321–344.

1984; Mandler, 1988; McDermott and Davis, 1984; McNamara, 1992; Montello, 1992, in press). Some models proposed for human spatial knowledge of the environment have suggested that it is best described as metric. A metric geometry describes spaces that have properties such as symmetry and the triangle inequality, properties that define quantitative measurement on spatial dimensions (see Montello, 1992; Shepard, 1964). Others have suggested that spatial knowledge is characterized by a less sophisticated geometry than a metric geometry. Geometries that are less than metric (e.g., topologies) do not define such quantitative properties, but include qualitative properties such as connectivity and containment. Still others have proposed compromise models that include multiple knowledge stores, one or more that is metric and one or more that is nonmetric.

### *Qualitative Metrics*

Within the past decade, there has been considerable work within the AI (artificial intelligence) community to develop a formal model of spatial reasoning that can reason well without the necessity of very precise metric knowledge or elaborate decision algorithms. This work has taken place within the larger context of *qualitative reasoning*, a term describing models that reason fairly effectively about a variety of problems without sophisticated and precise calculation abilities. Work in *naive* or *qualitative physics*, for example, has attempted to predict the motion of pulley systems without the use of precise physical data and rules of calculus (e.g., Forbus et al., 1991). In turn, qualitative reasoning models have derived much of their inspiration from the now robust topic of *fuzzy logic* (Dutta, 1990; McDermott and Davis, 1984; Zadeh, 1975).

For instance, Dutta (1988) provides a fuzzy model of spatial knowledge in which a statement about distance and direction is modeled as two fuzzy categories, each category consisting of a center value, and left and right intervals of spread. The statement "object A is about 5 miles away", for example, is modeled as having a center of 5 miles and 1 mile spreads around 5 miles. The statement essentially says that the distance is between 4 and 6 miles. The statement "object A is in about a north-easterly direction" is modeled as having a center at 45° and 10° spreads around 45°. The statement essentially says that the direction is between 35° and 55°. In both cases, the correct value is modeled as having some nonzero probability of falling within the category spreads. As we will discuss below, however, modelers such as Dutta provide no a priori reasoning or empirical evidence as to deciding how large this spread should be.

Such imprecise and inelaborate models of reasoning about spatial quantities have been dubbed *qualitative metrics*. They hold promise as models of human environmental spatial knowledge. Presumably, the models could account for human abilities and limitations at skills such as navigation and communication about space. Qualitative modelers have noted several difficulties with information processing in the real world, including perceptual imprecision, temporal and memory limitations, the availability of only approximate or incomplete knowledge, and the need for rapid decision-making (Dutta, 1988, 1990). One of the attractive properties of such approaches is possibly providing a way to incorporate both the metric skills and metric limitations of human spatial behavior without positing separate metric and topological knowledge structures.



Most of the work on qualitative metrics has focused on knowledge of directions in the environment necessary for navigation and spatial communication. Although the details of these proposals vary, they agree in positing a model of directions which consists of a small number of coarse angular categories, commonly four 90° categories (front, back, left, right) or eight 45° categories (front, back, left, right, and the four intermediate). Frank (1991a, 1991b) provides good examples of such approaches. His models consist of either 4 or 8 "cones" or "half-planes" of direction. Values along the category boundaries are considered "too close to call" and result in no decision about direction. He also provides a set of operators for manipulating these values. Other writers provide similar models of directional knowledge (Freksa, 1992; Hernández, 1991; Ligozat, 1993; Zimmermann, 1993).

It must be noted that AI researchers in general, and qualitative spatial modelers in particular, are not motivated exclusively or even primarily by a desire to simulate human knowledge and behavior accurately. In many cases, they may simply wish to design an intelligent system that works. Such an approach may only implicitly or incidentally produce a model of human spatial thought, if at all. But it should also be stressed that qualitative metric modelers have definitely taken inspiration from what they consider a realistic approach to human reasoning about space (and time):

> "It is a truism that much of human reasoning is approximate in nature. Spatial reasoning is an area where humans consistently reason approximately with demon-strably good results." (Dutta, 1988: 126).

> "Spatial reasoning is ubiquitous in human problem solving. Significantly, many aspects of it appear to be qualitative" (Forbus et al., 1991: 417)

> "Much of the knowledge about time and space is qualitative in nature. Specifically, this is true for visual knowledge about space." (Freksa, 1991: 365)

> "Our goal is to establish qualitative spatial relations between objects in a cognitively plausible way." (Hernández, 1991: 374)

> "a new approach is presented...to combine knowledge about distances and positions in a qualitative way. It is based on perceptual and cognitive considerations about the capabilities of humans navigating within their environments." (Zimmermann, 1993, 69).

In spite of all of this apparent wisdom about human spatial reasoning, these modelers have nearly completely avoided citing any behavioral research to support such conclusions. Our purpose in the work reported below is to make an initial attempt to evaluate qualitative metric models against empirical data from human subjects. The empirical data consists of estimates by humans of the sizes of turns of pathways they have walked.

## Empirical Data and Simulation Approach

A two-stage approach was taken in order to evaluate qualitative metric models of human directional knowledge empirically. In Stage 1, existing models from the qualitative reasoning literature were used to develop testable simulations that were maximally faithful to those models. The results of Stage 1 were used to design improved simulations in Stage 2. These Stage 2 simulations went beyond the existing qualitative metric models, but nevertheless attempted to retain the fundamental insight of a qualitative metric. That insight is the proposal that humans employ a small, finite number of quantitative categories to organize spatial knowledge. In both stages, Monte Carlo simulations were carried out in which estimates of turns were generated by randomly sampling within the discrete categories suggested by the model.

Some writers have proposed models in which entire single categories constitute responses (e.g., a forward response is a cone with a range of 90°); the precision of the response is not greater than the entire category (see Freksa, 1992; Zimmermann, 1993). Such a model is incompatible with the requirement of many behavioral studies, including the study we compare to our simulations in the present research, for subjects to estimate at much higher levels of precision. Of course, the fact that people readily provide estimates at the level of precision of one or a few degrees does not ensure that their knowledge is stored or is accurate at that level of precision, but it does suggest that similar precision should be designed into the simulations in order to ensure comparability. Therefore, random responding at the level of the single degree within discrete categories was considered the most realistic way to model the qualitative metrics in the simulations below.

The data used to evaluate qualitative metric models of directional knowledge came from the results of some research by Sadalla and Montello (1989). This research was an investigation of subjects' knowledge of turn sizes after walking pathways containing a single turn. Vision-restricted subjects who could see only the floor down around their feet walked an 8.3 m pathway marked on the floor containing two straight segments and one turn. There was a .5 m gap between the end of the first segment and the beginning of the second segment (thus not providing a completed visible angle). On different trials, the



size of the turn from straight ahead was varied. All subjects walked and estimated 11 different turns in different random orders, ranging in size from 15° to 165° from straight ahead and separated by 15° increments (Figure 1). Thus, the least extreme turn from straight ahead is labeled *15°* and the most extreme turn is labeled *165°*. Half of the subjects walked turns to the right, the other half to the left. This variable was unrelated to estimation performance and was not considered further by Sadalla and Montello, nor is it considered below.

After walking to the end of the path, subjects used a circular pointer to provide three separate measures of their knowledge of the angular size of the pathway turn. Two of these measures are used below to evaluate the simulations. Measure 1, henceforth called

Figure 1

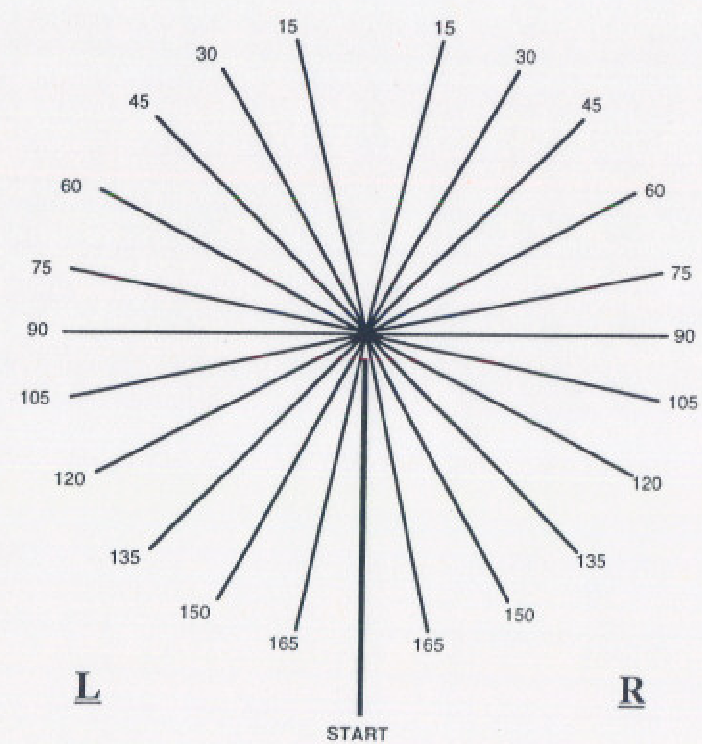



**Figure 1:** Angular pathways walked by subjects in Sadalla and Montello (1989).

*Turn Size*, required subjects to "reproduce the size of the turn"; measure 2, called *Original Direction*, required subjects to "point in the original direction of travel". Thus, correct answers to the two measures would be mirror images for any given turn. The latter measure is expected to result in poorer performance, however, because it compounds error from two sources: estimation of the angle of the turn size and estimation of the angle of the original direction from the heading direction at the end of the path. Measure 3 from Sadalla and Montello required subjects to "point back to the start location" and is not used here to evaluate the simulations because it involves both distance and directional knowledge.

Our thinking about human knowledge of directions in the environment is guided by a simple psychological process model that describes the acquisition, storage, and use of that knowledge (e.g., Simon, 1979). The model consists of five stages: (1) perception, (2) encoding, (3) long-term memory storage, (4) retrieval and recoding in working memory, and (5) behavioral output. Information about directions is perceived from the environment or from body movement. This information is encoded and stored in long-term memory. When the information is needed (e.g., for wayfinding decisions or researchers' requests), it is retrieved from long-term memory and placed in working memory (short-term memory). Various recoding processes (e.g., scale construction, image manipulation, verbalization) are brought to bear on the working-memory representations in order to produce behavioral outputs such as turn reproductions.

At various stages in the model, processes occur which produce error in the directional knowledge. These error processes result in both inaccuracies and imprecision in knowledge, and in behavioral output. The processes are of three types: systematic bias, categorization, and random fluctuation. Systematic biases lead to reliable inaccuracies, as when a turn is repeatedly recalled as being closer to 90° than it actually was. Biases are thought to operate during the encoding or working-memory recoding stages, or both. Categorization leads to the imprecision of angular knowledge that characterizes qualitative reasoning. Categorization is essentially a process whereby continuous information is coded into discrete intervals. It is also thought to operate during the encoding or working-memory stages, or both. Finally, random fluctuations lead to inaccuracies that are not, however, reliable or repeatable. They are expected to operate at all stages of the model.

## Simulation Stage 1

Perhaps the most critical issue in designing the simulation models concerns the number of quantitative categories of knowledge to incorporate, essentially the level of precision expressed by the model. As discussed further below, existing proposals have not decided this issue in any empirically principled way. Rather, they have generally attempted to



produce adaptive reasoning with as few categories as possible. Two levels of precision ( directional knowledge were tested in the Stage 1 simulations, a 4-cone (cones of 90 each) and an 8-cone model (cones of 45° each) (Figure 2). Because the idea of qualitative metric is that the reasoner has no reliable information more precise than at th level of the spatial category, the directional cones were modeled as uniform randor distributions during Stage 1 (as opposed to normal distributions, for instance).

Figure 2

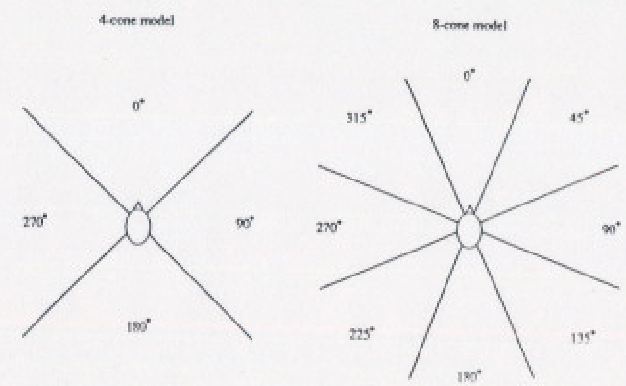



**Figure 2:** Homogeneous 4-cone and 8-cone models tested in Stage 1.

A second important issue in designing the simulation models concerns the accuracy c knowledge. Unfortunately, existing models do not address the accuracy issue; the assume perfect accuracy within their limits of precision. That is, existing models do nc incorporate any systematic biases in knowledge. All estimates for any turn falling within particular cone will be sampled from that cone. This approach to accuracy is modeled i Stage 1 in two ways. First, all estimates for a given pathway turn are consistentl sampled from within a single correct cone, referred to as *single-cone* sampling (turn falling on cone boundaries were considered ambiguous, however, and were sample equally from the two neighboring cones). Single-cone sampling will result in constan errors that vary across turns (imperfect accuracy). Sampling during Stage 1 was done in second way by proportionately sampling from two neighboring cones so that no averag inaccuracy resulted for any turns, referred to as *proportional* sampling. Proportiona sampling results in patterns of no constant error, perfect accuracy across turns (withi limits of sampling error). Each simulation was run enough times to generate 20( estimates for each turn.

Both the single-cone and proportional sampling approaches to the question of accuracy have implications for variability of performance as well. Of course, such perfect accuracy, even if imprecise, is almost certainly a poor model of human spatial knowledge. This is addressed directly in the Stage 2 simulations.

### *Results of Stage 1*

In order to evaluate the responses of the simulations, patterns and magnitudes of both constant error and variability from the simulations are compared to those from the empirical data set. Circular statistics (Batschelet, 1981) is used to calculate mean directions and mean angular deviations. These statistical techniques are appropriate for use with variables that consist of directional responses in 360°, called *circular variables* (or any periodic variable that shows cyclical trends). The techniques allow calculation of a mean angle or direction. Subtracting the mean direction from the correct answer provides a measure of *constant error* (systematic bias in one direction or the other). *Mean angular deviation* is also calculated as a measure of between-case variability (sometimes called variable error) in performance. It is the angular analogue to standard deviation. Mean angular deviation equals 0 when all directional estimates are exactly the same, and it reaches a maximum at just over 80° when directional estimates are maximally distributed around 360° (i.e., no agreement between subjects).

Constant errors for the empirical data (right and left turns collapsed) are depicted in Figure 3, both for Turn Size and Original Direction. Figure 4 depicts constant error for all four Stage 1 simulations: 4-single, 8-single, 4-proportional, and 8-proportional sampling. In all graphs of constant error, distortions toward 90° are graphed as positive errors, those away from 90° as negative errors.

Figure 3

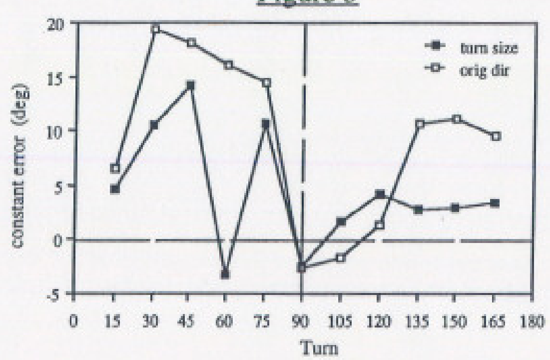



**Figure 3:** Constant error for empirical data from Sadalla and Montello (1989). Positive errors are distorted towards 90°.



Figure 4

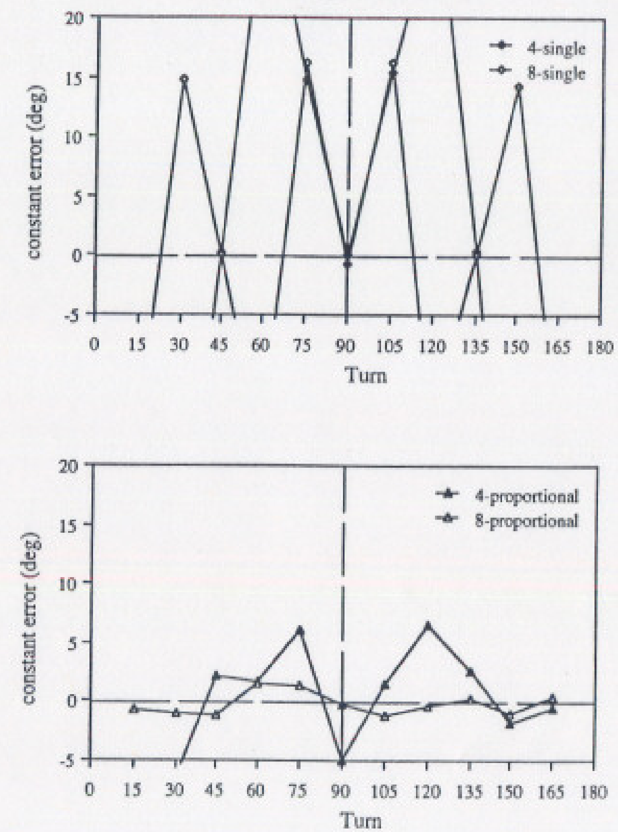



**Figure 4:** Constant error for Stage 1 simulation data. Positive errors are distorted towards 90° .

Constant error in the empirical data shows a distinctive pattern. There is a clear tendency for subjects to distort their turn estimates toward 90° (positive errors), with the error for Turn Size at 120° a clear exception that is likely due to a mistake in the original data collection. The range of distortion across turns is about 15°, with more distortion for acute than for obtuse turns (*acute* turns are 15-75°, *obtuse* turns are 105- 165°). The distortion is greater for Original Direction than for Turn Size. As discussed above, this is expected insofar as the former measure compounds the angular processing required for the latter measure.

None of the simulations mimic this pattern of accuracy well; most notably, none show a consistent bias toward right angles. It is true that the range of distortion is fairly well reproduced by the 4-proportional model, but this results from chance sampling error only. The 8-proportional model is nearly flat, showing very little distortion at all. Both single-cone models show a range of distortion across turns that is far too extreme, about 60° for the 4-cone model and 30° for the 8-cone model.

Figure 5 depicts mean angular deviations (variability) for both empirical measures. The corresponding results for the four simulations are depicted in Figure 6. Variability in the empirical data shows a distinctive pattern that was in fact the focus of the original analysis by Sadalla and Montello (1989). There is low variability for turns at or near the orthogonal axes (0°, 90°, 180°), with gradually increasing variability towards oblique turns (45°, 135°). Average variability is about 30°, with a range across turns of about 20°. As with constant error, there is more variability for acute than for obtuse turns. Also like the constant error, there is greater variability for Original Direction than for Turn Size, at least in the acute quadrant.

Figure 5

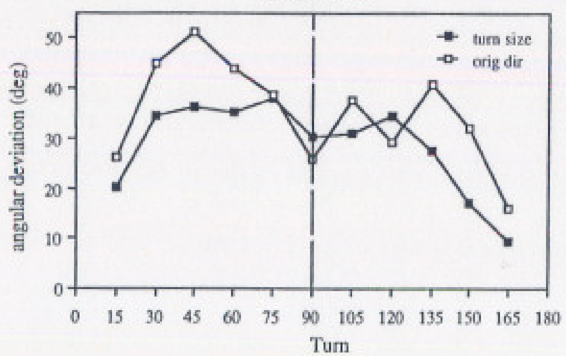



**Figure 5:** Mean angular deviation (variability) for empirical data from Sadalla and Montello (1989).

Interestingly, this pattern is reproduced to some extent by the 4-cone models: both proportional and single-cone sampling show the high agreement near orthogonal axes and the low agreement at oblique turns. The 4-proportional sampling results in a gradual increase to the obliques, while the 4-single sampling is flat for turns not exactly at 45° or 135°. But both 4-cone models produce overall variability that is too high (45-55°) and a range across turns that is too severe (30°). The 8- single model, on the other hand, results in a magnitude of variability (about 20°) that is too small in comparison to the empirical data and no change in variability across turns at all. Only the 8-proportional model results



Figure 6

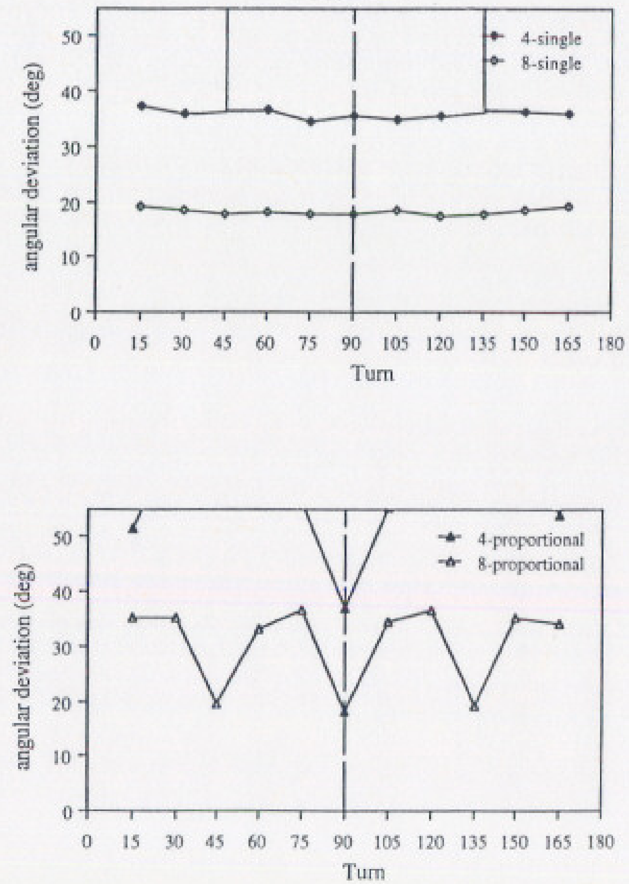



**Figure 6:** Mean angular deviation (variability) for Stage 1 simulation data.

in a magnitude of variability that matches the empirical data very well. This model results in an overall variability of about 25-30°, almost the same as the empirical data, and a range across turns of about 15°. However, neither 8-cone model reproduces the distinctively shaped pattern of the empirical variability. The 8-single model is flat, as mentioned above, showing no change in variability across turns at all. And the 8-proportional model results in lower variability at both 90° and the obliques, 45° and 135°. This drop at the obliques is the opposite of what is found in the empirical data. In addition, this model does not produce a decline in variability near 0° and 180°.

As a final approach to evaluating the performance of the simulations, circular correlations (Jammalamadaka and Sarma, 1988) are calculated between the correct turn values and both the empirical estimates and the simulated estimates. Circular correlation provides a measure of relationship between two circular variables. These are calculated within-case and averaged across cases using Fisher's r-to-z transformation to calculate mean correlations. The empirical estimates correlate very highly with the correct values, .94 for Turn Size and .89 for Original Direction. Both 8-cone models are similarly highly correlated with the correct values, .95 for single and .94 for proportional sampling. The 4-cone models also correlate strongly with the correct values but less so, .81 for single and .71 for proportional sampling.

### *Discussion of Stage 1*

Our Stage 1 attempt to evaluate qualitative metric models empirically suggests their possible viability as models of human directional knowledge. A qualitative metric model consisting of 8 45°-cones sampled proportionally reproduces the magnitudes of variability quite well. And both 8-cone models produce estimates that correlate with the actual turn sizes almost exactly as strongly as do the empirical estimates. The 4-cone models, on the other hand, produce far too much variability in performance and did not correlate with the actual turn sizes strongly enough.

Unfortunately, the 8-proportional model failed to reproduce the distinctive and oft replicated empirical pattern of minimal variability near orthogonal turns and maximal variability at oblique turns. Nor did it reproduce the empirical pattern of minimal constant error near orthogonal turns and distortion of turns toward 90° (empirical evidence of these patterns is cited and presented in Franklin et al., under review; Loftus, 1978; Sadalla and Montello, 1989; Tversky, 1992). These failures probably stem to a large extent from the lack of a reasonable approach to knowledge accuracy in existing qualitative metric models (we discuss this further at the end of the chapter). In our Stage 2 simulations, therefore, we attempt to improve the fit of the simulation results primarily by incorporating into the models some empirical and theoretical ideas about knowledge accuracy.

## Simulation Stage 2

The Stage 1 simulations were attempts to test qualitative metric models as they are currently described in the literature. Our approach in the Stage 2 simulations is to design more promising qualitative metric models of human spatial knowledge using the Stage 1 results as a guide. Because the 8-proportional model from Stage 1 did a good job of replicating the empirical magnitudes of variability and the within-case correlations with the actual turn sizes, it was decided to use 8-proportional approaches as a basis for our Stage 2 simulations.



Our discussion of the Stage 1 results above suggests several ways to modify and improve our initial simulations. Our first modification is to use heterogeneous cone sizes in an attempt to produce maximal variability near oblique turns and minimal variability near orthogonal turns. This is done in two ways: orthogonal cones of 30° and oblique cones of 60°, or orthogonal cones of 20° and oblique cones of 70°. In addition, we tried to decrease the variability for turns near 0° and 180° even more by splitting these two cones in half, producing cone sizes of 15°- 60°-30°-60°-15° or 10°-70°-20°-70°-10° (these are actually 10- cone models). These models are depicted in Figure 7. Decreasing the sizes of cones directly in front and behind the body by splitting them in half is consistent with the theoretical primacy of the front-back over the left-right axis of egocentric space (Franklin et al., under review; Shepard and Hurwitz, 1984) and the resulting maximal acuity of directional judgments near 0° and 180°. Franklin and Tversky (1990), for example, in a discussion of their *spatial framework*, found that subjects responded faster to locational queries about objects located in front or in back than about objects located to the left or right. The primacy is seen in our empirical data by the fact that variability for the 90° turn is generally greater than that for the 15°.

Figure 7

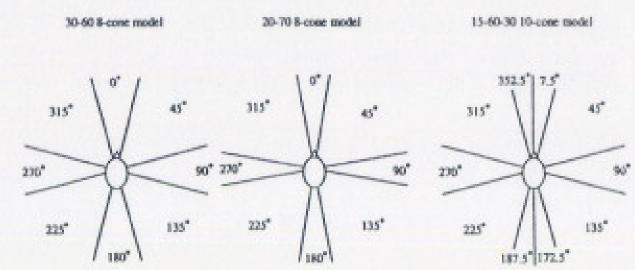



**Figure 7:** Heterogeneous 8 and 10-cone models tested in Stage 2.

Our second modification is designed to increase variability, particularly for turns near and at 45° and 135°. This "variability adjustment" is done by modifying the sampling of the 45°, 135°, and 90° turns. For all three turns, 10 cases are shifted from the cone containing the actual turn value to each of the two neighboring cones on either side. This adjustment method includes the 90° turn because the Stage 1 simulations had suggested that sampling entirely within a 20° or 30° cone would not produce enough variability for the 90° turn. Such an adjustment is also consonant with the lesser acuity for the left-right vs. front-back dimensions of space mentioned above. Each simulation was again run enough times to generate 200 cases.

Our third modification is to build in a *right-angle heuristic*. Any turn that deviates from straight ahead or straight behind is judged to be more nearly a right-angle turn than is actually the case. This is done by oversampling cones towards 90° for all turns other than the 90° turn, and is over and above the variability adjustments just described. Such a heuristic should produce the characteristic empirical pattern of distortion towards 90°. The heuristic is implemented in two ways: a fixed number of estimates are shifted one cone towards the 90° cone, or a percentage of estimates are shifted to the 90° cone. In either case, this adds sampling of the 90° cone for those turns that did not already sample it (such as the 15° turn).

In addition, the right-angle heuristic is made asymmetric by oversampling to a greater extent for acute than for obtuse turns. In the fixed number method, 10 cases are shifted toward 90° for the 15° and 165° turns, 40 cases are shifted for all other acute turns, and 20 cases are shifted for all other obtuse turns. In the percentage method, 20% of cases are shifted to the 90° cone for acute turns, and 10% are shifted for the obtuse turns. Either method should produce greater distortion for acute turns than for obtuse turns, and may also produce greater variability for estimates of acute turns.

Finally, we also examined the effect of using a different distribution to sample within each cone. All cones were sampled according to uniform distributions in the Stage 1 simulations. For the Stage 2 simulations, we also try sampling according to a normal, or Gaussian, distribution. These normal distributions were designed so that ±2 standard deviations covered the ranges of the corresponding cone (e.g., 30° or 60°).

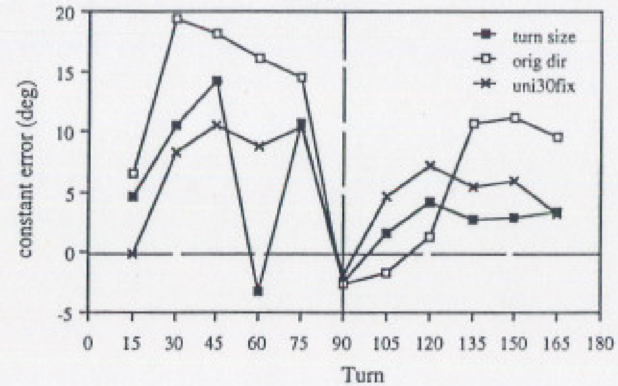
### *Results of Stage 2*

As was done with the Stage 1 simulations, patterns and magnitudes of both constant error and variability are compared to those from the empirical data set. All of the simulation variations described above were systematically varied across all possible combinations, but only the results from four simulations are considered here in detail. *Uni30fix* samples uniformly from 30° and 60° cones and biases according to a fixed number of cases. *Uni20fix* does the same but for 20° and 70° cones. *Uni15%* samples uniformly from 15°, 30°, and 60° cones (10-cone model), and biases according to a percentage of cases. *Norm15%* does the same but samples normally.

Constant errors for Uni30fix and Uni20fix are depicted in Figure 8, along with the empirical results for Turn Size and Original Direction. Constant errors for Uni15% and Norm15% are similarly graphed against the empirical results in Figure 9. Distortions toward 90° are again graphed as positive errors, those away from 90° as negative errors.



Figure 8



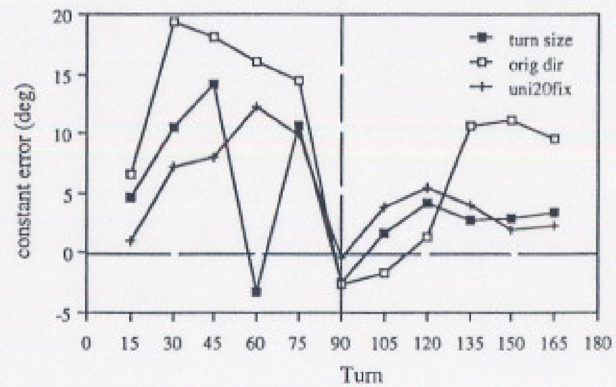




**Figure 8:** Constant error for Stage 2 simulation data (Uni30fix, Uni20fix) and empirical data. Positive errors are distorted towards 90°.

All four simulations recreate the empirical pattern of constant error quite well: minimal distortion near orthogonal turns, bias toward 90°, and greater bias for acute turns. The magnitudes of distortion match those for Turn Size very closely, with the exception of too little bias for the acute turns of 15-45°. Of note is the fact that the two alternatives for cone size are nearly equivalent; whether 30° and 60° or 20° and 70° cones are used makes very little difference (Figure 8). Similarly, whether uniform or normal distributions are used to model the cones makes very little difference (Figure 9). These conclusions were supported throughout all of the conducted simulations.

Figure 9

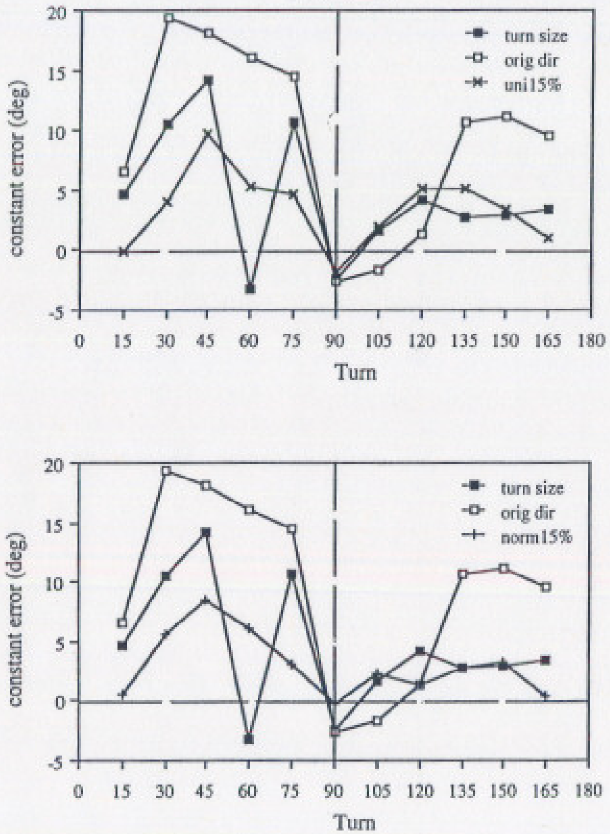



**Figure 9:** Constant error for Stage 2 simulation data (Uni15%, Norm15%) and empirical data. Positive errors are distorted towards 90°.

Similarly close fits are seen from an examination of variability in performance. Figure 10 depicts angular deviation for Uni30fix and Uni20fix, along with the empirical results for Turn Size and Original Direction. Figure 11 shows angular deviation for Uni15% and Norm15%. Again, all four simulations recreate the empirical pattern quite well: minimal variability near orthogonal turns, maximal variability for oblique turns, and greater variability for acute turns. The magnitudes of variability match those for Turn Size very closely. It can again be seen that the choice of cone sizes and sampling distributions is of little consequence, a result that was also echoed throughout all of the conducted



Figure 10

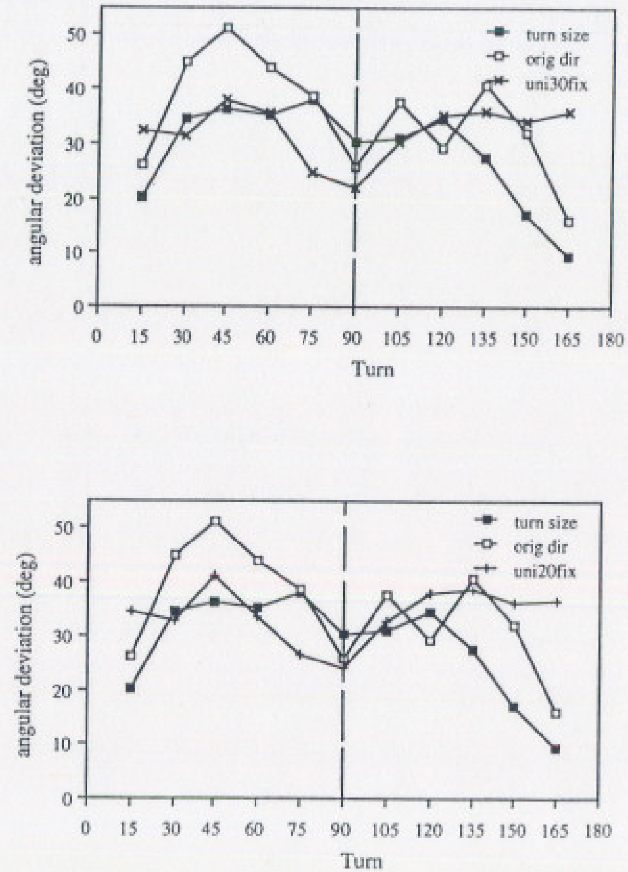



**Figure 10:** Mean angular deviation (variability) for Stage 2 simulation data (Uni30fix, Uni20fix) and empirical data.

simulations. One modification that does make a difference was splitting the 0° and 180° cones in half in the Uni15% and Norm15% simulations. This reduces the excessive variability for turns of 15° and 165° seen in the first two simulations, though still not to the level of the empirical data.

Finally, circular correlations between the correct turn values and the four sets of simulated estimates are calculated as in Stage 1. The mean correlations for Uni30fix, Uni20fix, and Uni15% are all .90, and that for Norm15% is .91. These compare

Figure 11

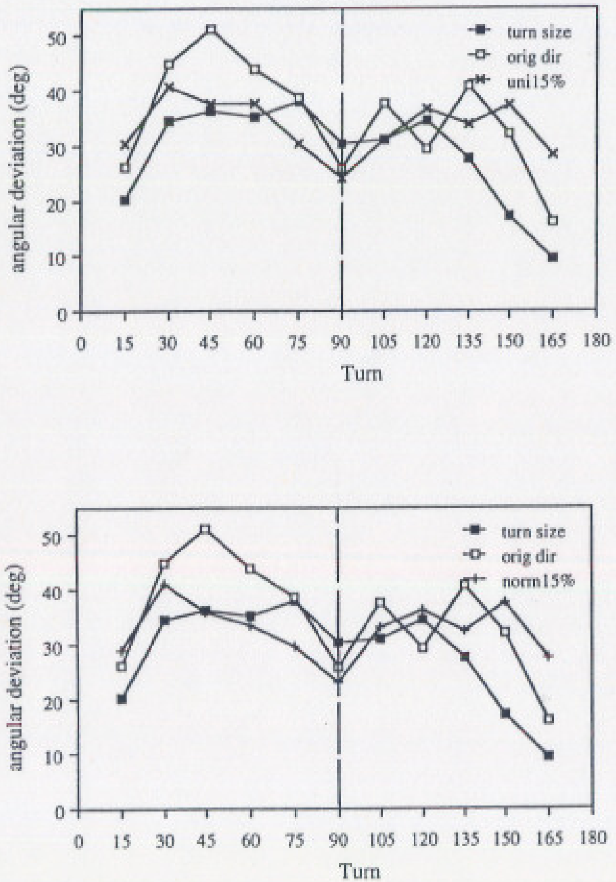



**Figure 11:** Mean angular deviation (variability) for Stage 2 simulation data (Uni15%, Norm15%) and empirical data.

favorably with the correlations of the empirical estimates with the correct values presented above, .94 for Turn Size and .89 for Original Direction.

*Discussion of Stage 2*

The Stage 2 simulations incorporated insights from the Stage 1 simulations and from other behavioral literature on directional knowledge. Their performance matched the



patterns of constant error and variability in the empirical data quite well, though not perfectly. This occurred without completely forgoing the essential insight of the qualitative metric literature. Each estimate was drawn randomly from one of a small number of cones, providing support for the idea that directional knowledge is imprecise and consists of a small number of quantitative categories of directions. Undoubtedly, other specific approaches to designing qualitative simulations could have been taken, though the relative insensitivity of the Stage 2 simulations to factors such as the exact cone size and nature of the sampling distributions suggests that additional variations in the specifics of the simulations would produce only modest improvements in their fit to the empirical data.

## Summary and Conclusions

The results of our simulations based on qualitative metric models provide some support for their promise as models of human spatial knowledge, but also suggest some of their serious weaknesses. Their major insight, that human knowledge is metric but only in a very imprecise way, receives support from these evaluations. In particular, some variety of proportional sampling within 8 or 10 directional cones produces magnitudes and patterns of constant error and variability that coincide closely with empirical patterns obtained from human-subjects testing. The exact sizes of the cones is apparently not critical, as long as orthogonal cones are smaller than oblique cones. It is also important to include in the model in some way a heuristic that oversamples cones towards 90°.

These simulations point to what is probably the most notable weakness of existing qualitative metric models, namely their lack of a reasonable approach to modeling knowledge accuracy. This conclusion derives from the empirical fact that none of the Stage 1 simulations, which were designed to be maximally faithful to existing models, produce anything like the pattern of bias towards right angles found in the empirical data used here or found in other research. Restricting sampling to single cones leads to unrealistically inaccurate performance; sampling proportionally from neighboring cones without implementing any biasing heuristic leads to unrealistically accurate performance. Also telling, however, is the difficulty and arbitrariness we experienced in trying to interpret existing proposals to guide how qualitative categories should be constructed and sampled. There was a definite ad hoc character to the decisions we made about how many quantitative categories to use, how large their respective ranges should be, and how they should be sampled. In its current state, therefore, the qualitative reasoning literature does not provide suitable models of human directional knowledge because it lacks an a priori principled approach to some of the central issues of such models.

As far as the Stage 2 simulations indicate, it is not important whether the cones are modeled as uniform or normal distributions. Models in which cones are described as consisting of central prototype values and normal variability around the prototype have been proposed in the behavioral literature (Huttenlocher et al., 1991). It is in fact a common approach to modeling fuzzy categories in the qualitative reasoning literature (e.g., Dutta, 1988, 1990; see also the discussion by McDermott and Davis, 1984). The distinction between uniform cones without internal structure and cones consisting of variability around a central value cannot be decided by these simulations.

The work in this chapter leads to ideas very similar to those found in Franklin and Tversky's (1990) discussion of their *spatial framework*. Notably, both research programs stress the idea of a Cartesian frame for organizing egocentric spatial knowledge. Two apparent contradictions with their work call for comment, however. One is the empirical finding by Sadalla and Montello (1989) that performance on acute turns (to the front) is more distorted and less precise than on obtuse turns (to the back). Franklin and Tversky not only found that subjects responded faster to front-back queries than to left-right queries, but also that responses to front queries were faster than to back queries. Franklin et al. (under review) replicated this and also found that directional pointing to objects in the front was more precise than to objects in the back across repeated trials (however, they also found that the range of directions which subjects would consider "to the front" was greater than to other directions). On the face of it, these results seem to contradict those by Sadalla and Montello. After all, trials with obtuse turns required subjects to turn around and walk "back", and to point in a backwards direction for both measures. And yet these trials resulted in lower variability than did performance after walking acute turns.

However, the tasks in the two sets of studies were different in some ways that may explain this apparent contradiction. Sadalla and Montello did not actually require subjects to point "to" anything behind their bodies; rather, these subjects had to move a pointer to show certain angular relationships on a circular pointer. The pointer had a movable wire and a radius drawn on its surface, thus providing a visible angle to subjects. Obtuse turns required the production of an acute angle on the pointing circle; acute angles are reproduced more precisely because of the stability provided by the context of the neighboring wire and radius line (cf. Pratt, 1926). Further, these subjects were responding with respect to a turn they had actually walked; subjects in Franklin et al. pointed to an object that they perceived visually. These considerations suggest that the quadrant asymmetries found by Sadalla and Montello may not necessarily be generally characteristic of egocentric directional knowledge. The issue calls for further clarification.

The second apparent contradiction with the *spatial framework* involves the superiority of at least an 8-cone model over a 4- cone model in the present simulations. The *spatial framework* essentially posits a 4-cone model, though not framed in that way (see especially, Franklin et al., under review). The empirical work by Franklin and her colleagues has not involved testing of qualitative metrics, however, and has not attempted to establish the number or sizes of cones necessary to produce appropriate estimation variability from random responding. And much of their work has concerned referents for natural language terms for direction ("front", "right", etc.) rather than nonlinguistic knowledge of directions. The two need not be synonymous.



The distinction between linguistic and nonlinguistic spatial knowledge is probably an important one to make. Several of the qualitative modelers in fact take inspiration from spatial information as expressed in natural language (Dutta, 1988, Fisher and Orf, 1991; Frank, 1991a, 1991b; Hernández, 1991; Zadeh, 1975; see also, Mark and Frank, 1991). Either they see language as equivalent to knowledge in general, or they have a specific interest in trying to model linguistic knowledge, for instance to improve natural language queries in Spatial Information Systems. However, there is a growing consensus that linguistic and nonlinguistic knowledge are not the same. The two involve at least partially separate systems, psychologically and physiologically. Furthermore, although linguistic knowledge of space is relatively limited in the precision of information it typically expresses (great metric precision is usually unnecessary), nonlinguistic knowledge may be much more metrically precise (Jackendoff and Landau, 1991; Landau and Jackendoff, 1993; McNamara, 1992; Rybash and Hoyer, 1992). Haber et al. (1993) found evidence for this difference directly relevant to our concern with directional knowledge. In their research, blind subjects indicated directions via several verbal and nonverbal methods; the verbal methods produced noticeably less precision than the nonverbal methods.

Linguistic issues aside, empirical studies could be conducted that would provide a more direct approach to designing and testing qualitative metric models than that provided by Sadalla and Montello (1989). The estimation variabilities from their research were actually obtained from between-subject performance. Within-subject variability is more to the point, however, because humans may organize knowledge with somewhat idiosyncratic quantitative categories. In an improved study, subjects would repeatedly walk each of a series of turns several times, providing multiple estimates of each turn. Furthermore, one could directly determine category ranges and boundaries by psychophysically establishing relative thresholds for angles and distances in a design that varies turn angles gradually over many repeated trials. A straightforward approach would require subjects simply to state which of two walked turns is greater or less than the other, or whether they are in fact equal.

Finally, it would be valuable to investigate qualitative models of distance knowledge to complement work on directional knowledge. Much less work on distance models has been reported in the qualitative literature. Frank (1991b) briefly discussed qualitatively modeling distance as consisting of two (near, far) or three (near, intermediate, far) categories; category size would depend on context. Fisher and Orf (1991) discuss a fuzzy set model of "near" and "close". Zimmermann (1993) exploits the information provided by two sets of half-planes connected by a directional vector. A triangulation between the two centers of the half-planes and a third point to be estimated results in some coarse knowledge about the distance of the third point relative to the two centers. In any case, it is likely that a successful qualitative model of distance would also incorporate heterogeneous category sizes (the smallest distance categories would have shorter ranges) and a biasing heuristic to produce overestimation of short distances relative to long distances (Montello, 1991).

In conclusion, this work leads to fruitful ideas about how to model human knowledge in a way that respects its metric qualities without imparting it with unrealistically excessive precision and accuracy. An interesting but as yet untested possibility is that the precision of human spatial knowledge decays over time (during long-term memory storage?). Spatial knowledge may exhibit rather precise metric qualities during perception (see Attneave and Pierce, 1978) and in memory after brief delays, delays characteristic of most empirical research. Over time, perhaps after delays on the order of months or years, there may be a continual degradation of spatial knowledge so that distortion and imprecision both increase. As one of us has suggested in the past (Montello, in press), such considerations support the need for rigorous very long-term longitudinal studies of environmental spatial knowledge.

## References

Attneave, F., and Pierce, C. R. (1978). Accuracy of extrapolating a pointer into perceived and imagined space, *American Journal of Psychology 91*, 371-387.

Batschelet, E. (1981). *Circular statistics in biology*, London: Academic.

Couclelis, H., Golledge, R.G., Gale, N. and Tobler, W. (1987), Exploring the anchor-point hypothesis of spatial cognition, *Journal of Environmental Psychology 7*, 99-122.

Dutta, S. (1988). Approximate spatial reasoning, In *First International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, pp. 126-140, Tullahoma, TN: ACM Press.

Dutta, S. (1990). Qualitative spatial reasoning: A semi-quantitative approach using fuzzy logic. In *Design and implementation of large spatial databases* (A. Buchmann, O. Gnther T. R. Smith, and Y. F. Wang, eds.), pp. 345-364. New York: Springer-Verlag.

Fisher, P.F. and Orf, T.M. (1991). An investigation of the meaning of near and close on a university campus. *Environment and Urban Systems 15*, 23-35.

Forbus, K. D., Nielsen, P. and Faltings, B. (1991). Qualitative spatial reasoning: The CLOCK project. *Artificial Intelligence 51*, 417-471.

Frank, A.U. (1991a). Qualitative spatial reasoning with cardinal directions. In *Proceedings of the Seventh Austrian Conference on Artificial Intelligence*, pp. 157-167. Springer-Verlag: Berlin.

Frank, A. U. (1991b). Qualitative spatial reasoning about cardinal directions. In *Autocarto 10* (D. Mark and D. White, eds.), pp. 148-167.



Franklin, N., Henkel, L. A. and Zangas, T. (under review). Parsing surrounding space into regions.

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

Freksa, C. (1991). Qualitative spatial reasoning. In *Cognitive and linguistic aspects of geographic space* (D.M. Mark and A.U. Frank, eds.), pp. 361-372. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Freksa, C. (1992). Using orientation information for qualitative spatial reasoning. In *Theories and methods of spatio-temporal reasoning in geographic space* (A. U. Frank I. Campari, and U. Formentini, eds.), pp. 162-178. Berlin: Springer-Verlag.

Golledge, R.G., and Hubert, L. J. (1982). Some comments on non- Euclidean mental maps. *Environment and Planning A 14*, 107-118.

Habe,r L., Haber, R. N., Penningroth, S. Novak, K. and Radgowski H. (1993). Comparison of nine methods of indicating the direction to objects: Data from blind adults. *Perception 22*, 35-47.

Hernández, D. (1991). Relative representaion of spatial knowledge: the Z-D case. In *Cognitive and linguistic aspects of geographic space* (D.M. Mark and A.U. Frank, eds.), pp. 373-386. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Huttenloche, J., Hedges, L.V. and Duncan S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review 98*, 352-376.

Jackendoff, R., and Landau, B. (1991). Spatial language and spatial cognition. In *Bridges between psychology and linguistics: A Swarthmore Festschrift for Lila Gleitman* (D.J. Napoli and J.A. Kegl, eds.), pp. 145-169. Hillsdale, NJ: Lawrence Erlbaum.

Jammalamadaka, S.R. and Sarma, Y.R. (1988). A correlation coefficient for angular variables. In *Statistical theory and data analysis II* (K. Matusita, ed.), pp. 349-364. North- Holland: Elsevier Science Publishers B.V.

Kuiper,s B.J. and Levitt, T.S. (1988). Navigation and mapping in large-scale space. *AI Magazine, summer*, 25-43.

Landau B., and Jackendoff R. (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences 16*, 217-265.

Landau, B., Spelke, E., and Gleitman, H. (1984). Spatial knowledge in a young blind child. *Cognition 16*, 225-260.

Ligozat, G.F. (1993). Qualitative triangulation for spatial reasoning. In *Spatial information theory: A theoretical basis for GIS* (A. U. Frank and I. Campari, eds.), pp. 54-68. Berlin: Springer-Verlag.

Loftus, G.R. (1978). Comprehending compass directions. *Memory & Cognition 6*, 416-422.

Mandler J.M. (1988). The development of spatial cognition: On topological and Euclidean representation. In *Spatial cognition: Brain bases and development* (J. Stiles-Davis, M. Kritchevsky, and U. Bellugi, eds.), pp. 423-432. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mark, D.M. and Frank A. U. (eds.). (1991). *Cognitive and linguistic aspects of geographic space.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

McDermott, D. and Davis, E. (1984). Planning routes through uncertain territory. *Artificial Intelligence 22*, 107-156.

McNamara, T.P. (1992). Spatial representation. *Geoforum 23*, 139- 150.

Montello D. R. (1991). The measurement of cognitive distance: Methods and construct validity. *Journal of Environmental Psychology 11*, 101-122.

Montello, D.R. (1992). The geometry of environmental knowledge. In *Theories and methods of spatio-temporal reasoning in geographic space* (A. U. Frank, I. Campari, and U. Formentini, eds.), pp. 136-152. Berlin: Springer-Verlag.

Montello, D.R. (in press). A new framework for understanding the acquisition of knowledge about large-scale environments. In *Sptial and Temporal Reasoning in Geographic Information Systems* (R.G. Golledge and M. Egenhofer, eds.).

Piage, J. and Inhelder, B. (1948/1967). *The child's conception of space*. New York: Norton.

Pratt, M.B. (1926). The visual estimation of angles. *Journal of Experimental Psychology 9,* 132-140.

Rybash, J.M. and Hoyer W. J. (1992). Hemispheric specialization for categorical and coordinate spatial representations: A reappraisal. *Memory & Cognition 20,* 271- 276.

Sadalla, E.K., and Montello, D.R. (1989). Remembering changes in direction. *Environment and Behavior 21*, 346-363.

Shepard, R.N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology 1*, 54-87.

Shepard, R.N. and Hurwitz, S. (1984). Upward direction, mental rotation, and discrimination of left and right turns in maps. *Cognition 18,* 161-193.

Siegel, A.W. and White, S.H. (1975). The development of spatial representations of large-scale environments. In *Advances in child development and behavior* (H. W. Reese, ed.), pp. 9-55. New York: Academic.

Simon, H.A. (1979). *Models of thought.* New Haven, CT: Yale University Press.

Tversky, B. (1992). Distortions in cognitive maps. *Geoforum 23,* 131-138.

Zadeh, L.A. (1975). Fuzzy logic and approximate reasoning. *Synthese 30*, 407-428.

Zimmermann, K. (1993). Enhancing qualitative spatial reasoning -- combining orientation and distance. In *Spatial information theory: A theoretical basis for GIS* (A. U. Frank and I. Campari, eds.), pp. 69-76. Berlin: Springer-Verlag.

Daniel R. Montello
Department of Geography
University of California, Santa Barbara

Andrew U. Frank
Department of Geo-Information
Technical University of Vienna