SPATIAL COGNITION AND COMPUTATION, *3*(2&3), 185–204

Where's Downtown?: Behavioral Methods for Determining Referents of Vague Spatial Queries

Daniel R. Montello, Michael F. Goodchild,
Jonathon Gottsegen, and Peter Fohl
University of California at Santa Barbara

Humans think and talk about regions and spatial relations imprecisely, in terms of vague concepts that are fuzzy or probabilistic (e.g., *downtown*, *near*). The functionality of geographic information systems will be increased if they can interpret vague queries. We discuss traditional and newer approaches to defining and modeling spatial queries. Most of the research on vague concepts in information systems has focussed on mathematical and computational implementation. To complement this, we discuss behavioral-science methods for determining the referents of vague spatial terms, particularly vague regions. We present a study of the empirical determination of *downtown Santa Barbara*. We conclude with a discussion of prospects and problems for integrating vague concepts into geographic information systems.

Keywords: Vagueness, Spatial Queries, Cognitive Regions, Geographic Information

People typically think and communicate about the world in terms of vague concepts. Unlike formal languages, natural languages used in everyday speaking and writing typically refer to categories that do not have precise referents and are not delimited by sharp semantic boundaries. Furthermore, unlike formal concepts such as those of geometry, exemplars of vague concepts vary in the degree to which they are members of a category or the chance that they are members of a category; that is, they are *fuzzy* or *probabilistic* (Lakoff, 1987; Smith & Medin, 1981; Zadeh, 1965). For example, Rosch and Mervis (1975) showed that lay people generally consider robins to be better examples of

---

Correspondence concerning this article should be addressed to Daniel R. Montello, Department of Geography and the NCGIA, The University of California, Santa Barbara, CA 93106, USA; email montello@geog.ucsb.edu

birds than are penguins, though both are birds to some degree, and both are birds according to the more rigorously-defined criteria of ornithology.

Natural language about space and place is no exception. Two classes of vague spatial terms are commonly used in geographic communication and thought: *spatial relations* and *regions*. Vague (qualitative) spatial relations include such terms as *near*, *around*, and *to the east* (Altman, 1994; Mark & Frank, 1989; Retz-Schmidt, 1988). Similarly, regions, which are essentially categories of land surface area, are typically vague (Mark & Csillag, 1989). *Administrative* regions such as a states or land parcels have sharp boundaries imposed on them (Smith & Varzi, 1997). But other region concepts used by lay people refer to probabilistically graded or fuzzy entities (as do the concepts of *thematic* and *functional* regions used in geo-science research contexts—see Montello, 2003). Examples of such *cognitive* or *perceptual* regions include *downtown*, *Riviera neighborhood* (in Santa Barbara, California), and *Midwest*. For our purposes, the two classes of spatial relations and regions share many similarities. Both refer to spatial extents without precise boundaries, and for which there are no exact criteria for membership—no finite set of necessary and sufficient characteristics. A formal test of the relationship "A is to the east of B," for example, might require that there exists at least one due east-west line that intersects both A and B. Informally, however, such a directional reference is likely to be used under a range of conditions that are difficult to identify precisely (Frank, 1996). Similarly there is no formally defined, universally accepted line that demarcates the Midwest, and any two individuals will agree only partially about which areas of the United States are part of the Midwest. Some areas (typically near the center of the region) are considered to be better or more typical examples of the Midwest than are other areas.

In this paper, we discuss the use of vague spatial concepts, particularly vague regions, in geospatial thought and communication. Given the ubiquitous use of vague spatial concepts, we agree with the premise (e.g., by Kuhn, 2001) that the functionality of geographic information systems (and other spatial information systems) will be enhanced if they can interpret queries containing vague terms. Our focus in this paper is on ways to determine the referents of queries about vague regions in geospatial information systems; what do people mean when they ask for a map of "Northern California" or the "area around the Eiffel Tower?" The importance of understanding vagueness has been widely recognized in geographic information science for at least a decade (e.g., contributions in Burrough and Frank, 1996) and even longer in other disciplines (e.g., Zadeh, 1965). There are many examples of work that discuss how to mathematically or computationally represent vagueness; solutions have included fuzzy logic, multivalued logic, probabilistic logic, and more (Altman, 1994; Cobb et al., 2000; Cohn & Gotts, 1996; De Bruin, 2000; Mark & Csillag, 1989; Papadias et al., 1999; Wahlster, 1989; Wang & Hall, 1996). Our focus in this paper is not on the formal structure of vague spatial concepts, though this work is obviously critically important to implementing vagueness in information systems. With few exceptions (discussed below), however, work on formally implementing vagueness does not discuss how to determine *what* should be implemented (this "behavioral omission" is discussed by Montello and Frank, 1996). This is especially problematic for vague regions, because nothing in the formal mechanics of representing the "Midwest" specifies what land area should be included—what the "content" of the region is. In the present paper, we address the problem of how behavioral methods can be used to determine what people mean when they use vague terms, particularly vague spatial terms[1]. To demonstrate these methods, we present a detailed example of the empirical determination of *downtown Santa Barbara*. The paper concludes with a return to the general question of the impacts of geographic information technologies on interactions among humans, technology, and the environment.

Precise and Vague Spatial Concepts in Practice

Traditionally, a complex set of arrangements has allowed formal and informal approaches to defining spatial concepts to coexist in relative harmony. These arrangements are being disrupted, however, by the rapid introduction of digital information-processing technologies into the geographic domain (Goodchild & Proctor, 1997). A digital system is inherently precise, and thus favors rigorously defined concepts. There has been much discussion over the extent to which such technologies bias, filter, or otherwise intrude on the interactions between humans and their environment (e.g., Pickles, 1995).

In-vehicle navigation systems provide an example of a GIS that would benefit from the capacity to handle vague spatial concepts. Systems that use natural-language interfaces now exist. Some research suggests that natural language provides a better medium for communicating spatial information in this context than does a strict reliance on maps (Streeter et al., 1985). Further research will attempt to determine the types of features and spatial relations that are most useful to include in computer-generated instructions, and how these features and relations should be verbalized (e.g., Allen, 1997).

Another example of a GIS that would benefit from the ability to handle vague spatial concepts is a digital map and imagery library (e.g., the Alexandria Digital Library at UCSB; Smith, Andresen, Carver, Dolin, Fischer, Frew, et al., 1996, and see http://www.alexandria.ucsb.edu). Users of conventional map and

[1] Many of the issues discussed in this paper also apply to a variety of other vague terms that are not explicitly spatial, including vague features, themes, and linguistic hedges, e.g., *pond*, *cold*, and *very* (e.g., Mark, 1993; Wallsten et al., 1986; Wang, 1994). An important distinction between explicitly spatial and nonspatial vague terms is that spatial terms involve delineation of portions of space as a literal entity, while nonspatial terms may be mapped onto space as a metaphorical entity (as in "semantic" space).

imagery libraries frequently pose queries based on vague regions, and a complex interaction between user and librarian is often needed before the query can be satisfied; furthermore, the result rarely meets the user's needs exactly. By contrast, the digital world is inherently precise, and approaches to queries are often essentially Boolean. For example, the Federal Geographic Data Committee's Content Standards for Digital Geospatial Metadata (http://www.fgdc.gov) include a number of fields, each corresponding to some defining characteristic of geographic data sets. The process of searching for data sets that fit defined needs is thus precise, since each data set either matches or does not match the specification of the search. It is difficult to incorporate the essential vagueness and trial-and-error of the conventional approach in a precise digital system.

In practice, geographic regions are more likely to be identified by their limits than by enumerating their contents. In response to a query, a digital library system must decide the extent of area to return to the user, and this is essentially the problem of identifying the limits, or boundaries, of the region in question. But there is a more conceptually interesting reason for a focus on boundaries. Because the geographic surface is continuous, a finite region will contain an infinite number of locations; thus, definition by enumeration is possible only if a region is defined as an aggregate of a finite number of smaller regions. Specification of a region is therefore often reduced to specification of its boundary. Even in the case of a well-defined region, however, specification of a boundary is ultimately insufficient to define a geographic region. Suppose, for example, that a boundary is defined as following a parallel of latitude. In principle, it would be possible to determine whether a point was inside or outside the region by measuring its latitude. Unfortunately, our ability to measure latitude is limited by the nature of our measuring procedures. Thus, the traditional response has been to replace a boundary defined by latitude with a series of physical monuments on the ground, together with the rule that the boundary between any adjacent pair of monuments follows a straight line. Additional uncertainty results when the monuments are lost, subject to tectonic movement, or represented digitally. In an important sense, then, all geographic regions can only be ultimately identified in a vague way, because the location of a point can never be measured accurately with unlimited precision.

These difficulties are even more profound for the vague spatial concepts we consider in this paper. They are vague in part because people are unsure about their precise referent— essentially a variant of measurement error, many spatial concepts reflect *epistemological* vagueness. But these concepts are vague in a fundamental *ontological* sense too, not just because of limitations due to measurement error, disagreements among experts, or inattention to temporal and scale issues (Burrough & Frank, 1996), though all of these are important reasons why GIS needs to deal better with vague information. They are cognitively vague, represented in people's conceptual understandings of the world as vague entities.

In many contexts, vague terms have either been replaced with precisely-defined ones, or ignored. Thus the adjectives *hot* and *cold* have been replaced by precise scales of temperature. Similarly, certain vague geographic regions have been replaced by precisely-defined regions by surveying their boundaries, by arbitrarily identifying precise boundaries, or by defining them as aggregates of well-defined component parts (e.g., biogeographic regions are discussed by Gray, 1997). But many regions lack official recognition, remain the subject of debate, or are tacitly accepted as being part of an informal geographic language. Gazetteers, used to describe the index of place-names found in atlases, have connotations of official recognition, typically including only the names of administrative entities that have some level of formal definition. The gazetteer of the Geographic Names Information System (GNIS) of the U.S. Geological Survey (http://mapping.usgs.gov/www/gnis/), a list of place-names derived from the USGS's topographic maps and arguably a digital equivalent of an atlas gazetteer, similarly reflects a preference for places with some form of official recognition, and omits less formal terms such as *Riviera* or *Midwest*, terms in common use and likely subjects of library searches.

Determining the Referents of Vague Spatial Queries

Behavioral-science methods are needed to determine the referents of vague queries. As we stated above, the various formal approaches to defining vague spatial concepts (fuzzy logic, etc.) may provide the computational mechanics for implementing vague concepts but do not provide a principled basis for determining the actual content of the implemented concept. This content is an empirical question, and requires empirical observation or interrogation of human conceptualizers. In the study we report here, we demonstrate a method for empirically determining from human subjects the referent of a vague region, *downtown Santa Barbara*.

Aside from the traditional solutions of ignoring vague queries or making them artificially precise, there are two approaches one might take to empirically determining their referents (Burrough, 1996; Robinson, 2000). The first is the *a priori* approach, in which an understanding of particular vague terms is stored in the system. For example, a representation of a commonly used but informal term, such as *downtown Santa Barbara*, could be stored in the system's database along with formally-defined terms, using appropriate representations. To implement this approach, one would need to conduct empirical tests or interviews with human informants and store the results. The second approach is *interactive*; no prior understanding of the vague term is stored in the system. Instead, the system interrogates users in some way to determine the referents of their vague queries. Conveniently, the methods useful for an a priori approach can be applied to an interactive approach. To do this, one or more of the data collection procedures we describe below would be implemented into the interface of the system as a series of system-generated queries. The system would process the responses to those queries in real time in order to make decisions about what areas of space to represent to users as the referents of their queries.

Robinson (1990, 2000) introduced such an interactive method for determining the referents of vague spatial relations like *near* and *far*; in doing so, he provides a rare example of an empirical attempt to determine the content of geospatial natural-language concepts (Mark, 1993, is another early example). Like our research, Robinson's work represents an attempt to help solve the problems of creating natural-language GIS. In his research, Robinson programmed a computer to ask a series of yes/no questions of a user, such as "is Alma near Douglas?" The program presents a series of such questions, based in part on the pattern of answers it gets. The research we present below may be seen in part as an attempt to replicate and extend Robinson's research, but focusing on vague regions rather than relations.

The empirical methods we propose to determine the referent of a user's vague spatial query are based on two assertions. The first is that a region like *downtown* is in fact vague rather than precise in respondents' conceptual structures. The second, shared with Robinson, is that respondents will be willing to make discrete judgments about area membership even though they believe that the areas do not have precise boundaries. In other words, we assume that users will typically be comfortable making "judgment calls" about the regional membership status of any small piece of the Earth. The responses we obtained in our study reported below support both of these assertions.

We conceive of a vague object as a field z(X), giving a measure of the object's presence at any point in the plane (or on the surface of the Earth). A well-defined object is a binary field, $z = \{0,1\}$. But for vague objects, various interpretations of z are possible. Blakemore (1984) and others have suggested that z be three-valued, with an intermediate value denoting "X may be in A." In the egg-yolk model of Cohn and Gotts (1996), the "yolk" is "in A," the "white" is "may be in A," and the two sets together form the "egg." Since z has only three values in this model, its spatial variation can be treated as a simple variant of the standard Boolean two-valued case. In the more general case, however, the scale of z is continuous. It could be interpreted as a probability, $0 \leq z \leq 1$, either strictly frequentist as the probability that a randomly chosen person or observer would assign the point to A, or subjectively as a measure of an observer's confidence in assigning the point to A. Alternatively, it could be interpreted as a measure of the membership of X in the fuzzy set A (Zadeh, 1965).

In principle, determining the field z(X) could require an infinite (or at least very large) amount of sampling, if the value of the field were independent at every location. In practice, however, vague spatial concepts tend to be strongly autocorrelated, so that effective determination may be possible with fairly sparse sampling. If we give z(X) a frequentist interpretation, a surface can be constructed by interrogating a sufficiently large sample of users. Each user would be asked to identify locations that lie within the region [z(X)=1] and locations that lie outside [z(X)=0]. Again, sampling would continue until sufficiently dense. After a sufficient number of users have provided input, z(X) would be computed by some appropriate method of convolution over the inputs. For example, if every user provides input at the same set of sample locations, z(X) can be obtained by a simple average at every point. If point sample locations do not coincide between users, z(X) could be obtained by a convolution such as:

$$z(X) = \sum_j x_j W_j \Big/ \sum_j W_j$$

where j is a sample point, $x_j=\{0,1\}$ the value assigned at that point, and W_j is a decreasing function of the distance d_j between x and point j, such as $\exp(-bd_j)$. We term the result a *frequentist* z(X).

Interpreting z(X) as a subjective probability or fuzzy membership, we could ask users to assign values of z directly to locations. For example, users could be asked to indicate locations that they are "50% confident" lie within the region, or to draw the location of the "50% confidence" isoline. The process could be continued until the entire area is sampled sufficiently densely for a representation of z(X) to be built. An appropriate method of spatial interpolation would then be used to determine z(X) at any remaining locations. We term this a *subjectivist* z(X).

The specific methods we use in this study employ both of these approaches. The first method is distinctly frequentist in concept and probably the most efficient in placing minimal demands on the user. Presented with an inclusive base map, a user simply draws a line around the area believed to constitute the referent of the query (Figure 1a). Aitken and Prosser (1990) employed this method to determine residents' beliefs about the extents of their neighborhoods. Similarly, Brown (1991) had respondents identify downtown Tacoma, Washington, by verbally describing the boundaries or by marking the boundaries on a street map overlaid on an aerial photograph of the area. A frequentist z(X) can be obtained by averaging the binary surfaces obtained in this way from a number of users.

Other designs might be used to elicit similar binary information. For example, we could present the user with a series of sample points or raster cells, asking in each case for a binary response—the point or cell lies inside the region, or outside the region (Figure 1b). Points could be presented on a regular grid, with a spacing determined by the resolution needed and by the ergonomics of the task. But clearly it would be more efficient to concentrate sample points in the region of the boundary, where the variation of the frequentist z(X) is highest (Figure 1c). This is essentially the method Robinson (1990) used to elicit meanings of *near* and *far*. While complex algorithms might be devised to selectively sample the boundary region, this is effectively what happens when

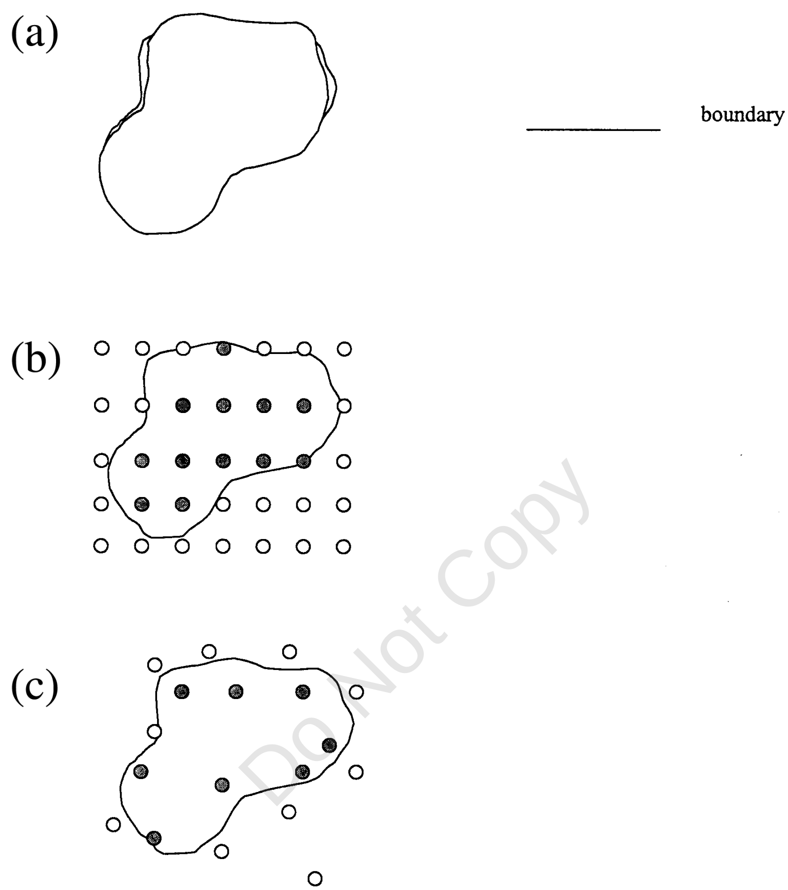

Figure 1. Strategies for eliciting individual representations of regions: a) by sketching a boundary; b) by responses over a grid; c) by selective trial-and-error sampling.

the user is asked to draw the actual boundary, and we conclude that the strategy of asking users to draw the boundary samples the plane very efficiently.

From a subjectivist perspective, the binary nature of the surface elicited from an individual respondent using the methods just described reflects a precise view of a region rather than a vague view. So as an alternative, various subjectivist methods might be devised that would interrogate the respondent for subjective probabilities or vague memberships at sample locations. We examine one of these below by asking respondents to draw the boundaries that they are 50% and 100% confident delimit downtown.

The base map. A technical problem arises in all of these methods because of the need to display a base map before eliciting representations of a region. To elicit an estimate of the region *downtown Santa Barbara*, for example, it is necessary to display and sample an area that includes the entire region. But until the region has been estimated, it is not possible to know whether the area displayed is sufficient. The interpretation of a region term will likely depend somewhat on the context of the user's specific problem; for instance, the location of downtown depends somewhat on whether one is thinking of shopping, dining, or attending the theater (though locations of these functions are clustered in space). Moreover, it is likely that a user's estimate will depend to a degree on the area shown; he or she will be more likely to underestimate a region that is large relative to the displayed area, and to overestimate a region that is small. We propose the following strategies to deal with these considerations:

1. Request that the user define the display area, by manipulating a comprehensive base map (panning, zooming, and clipping) until satisfied.

2. Request that the user delimit the boundary on the display. If the boundary approaches within a threshold distance of the edge at any time, interrupt and ask the user if the display should be redefined. If yes, return to 1. The threshold distance could be set to some proportion of the smaller of the linear dimensions of the screen, e.g., 0.15.

Empirical Study of Downtown Santa Barbara

To investigate the empirical determination of referents of vague regions, we conducted a study on people's beliefs about the extent of *downtown Santa Barbara*. We addressed several issues in this study: How well does a method for measuring vague regions work? What is the nature of the region that results? How do people understand and respond to instructions to draw boundaries of varying confidence? How feasible would it be to implement such regions in a GIS?

Method

Participants. Participants were pedestrians stopped at one of eleven locations on sidewalks near or within the area we anticipated that most people would consider downtown. These locations were chosen out of convenience as places with an intermediate amount of foot traffic. Potential participants were stopped at random at these locations. Our request informed them that we were "doing a research project on the way people identify neighborhoods," and that we wanted to ask "a few questions about the area that you consider to be downtown Santa Barbara." Thirty-nine people agreed to participate, 17 females and 22 males. One female and two males were excluded from data analysis because they did not complete the task. The remaining 36 participants ranged in age from 20 to

70 years, with a median age of 42. All except one lived and/or worked in the South Coast area of Santa Barbara County, which includes the city of Santa Barbara and the surrounding communities. Nearly half lived and/or worked in or near what they identified as the downtown area.

Materials. A base map showing the entire City of Santa Barbara and surrounding area on all sides was used for the mapping tasks. The base map was printed on American legal-size paper (8.5 x 14 in [21.6 x 35.6 cm]). The administrative boundary of the City was not visible on the map. A reduced version of the base map can be seen in Figures 2-5.

Procedures. Data collection involved three simple mapping tasks. The first task elicited participants' beliefs about the size and shape of downtown Santa Barbara. Participants were told to draw a line on the base map to "outline the area of the city that you consider to be downtown." Pilot tests had shown that some people would not delineate a closed area in response to this request, given that the city of Santa Barbara borders an ocean (we assume they meant us to consider their downtowns as ending at the waterfront). Closed areas were considered necessary for aggregating and displaying the vague regions, so for the regular data collection, we explicitly instructed participants to "enclose an area [around downtown] on the map." We refer to the regions generated from this first task as *default* regions, because they were elicited before we discussed the concept of vagueness with participants. They reflect a frequentist conception of vague regions, aggregated over participants.

The second task attempted to explicitly capture the subjectivist vagueness of an individual’s conception of downtown. We introduced the second task to participants by explaining that downtown is not formally defined, that there is no official boundary for downtown. We explained that they might feel that some places are more representative, more clearly part, of downtown than others. We checked to make sure that participants understood what we meant by this; in fact, nearly all participants understood this readily and expressed agreement with it. The map with the default polygon drawn on it was removed and a new base map was given to participants. We instructed them to draw a line on the new base map to show the area that they were "100% confident was downtown," pointing out that this might or might not be different than their first line. We then asked participants to draw, on the same map, the area that they were "50% confident was downtown." Thus, the second task generated *100% regions* and *50% regions*. For convenience, either might be termed *confidence regions*.

The third and final task captured participants' conceptions of the most representative point in downtown. Participants put an x at the location that they considered to be the “core or focal point of downtown." We carefully avoided referring to the *center* of downtown because we did not necessarily want to elicit their conceptions of downtown's geometric centroid. Rather, we were interested in their assessment of the point that most represents downtown, that was most clearly *in* downtown. Finally, we collected basic information about how long participants had lived in the Santa Barbara area, and the neighborhoods in which they lived and worked.

Results and Discussion

The raw-data polygons for each participant's default region are displayed in Figure 2. These are a set of overlapping polygons, the smaller being mostly contained within the larger ones. Figure 2 makes evident that *downtown Santa Barbara* is a vague region, when that is understood in the frequentist sense (i.e., participants drew nonidentical regions). However, the large degree of overlap, particularly around a central core, suggests that the vague regions of different participants are similar enough so that an attempt to measure and display a single region is a meaningful exercise.

The polygons were digitized and entered into a GIS. Figure 3 is a representation of these data in the aggregate, mapped using dot-density shading (Lavin, 1986). In order to produce these, the continuous function describing goodness of fit is approximated using discrete raster representations of the vague regions. The resulting map of downtown apparently communicates vague boundaries effectively, though this method produces some noncontinuous gradations near the periphery that are artifacts of sparse data in the boundary areas of the largest raw data polygons.

After the default regions were drawn, participants were asked to show 100% and 50% regions on a new base map. The 100% and 50% raw-data polygons are shown in Figure 4. Aggregated dot-density maps for both confidence regions are shown in Figure 5.

It appears that the instruction to draw confidence regions of varying degrees of confidence (as expressed in percentages) was interpretable in sensible ways by most participants. Of the 36 participants, 33 drew 50% regions that are larger than their 100% regions, the other three drawing them as equal. Furthermore, all but 2 participants drew 50% regions that wholly contained their 100% regions.

It is interesting to compare the default regions to the two confidence regions because it suggests how participants interpreted the default instructions. Nearly all participants drew different polygons in response to the three tasks. Specifically, only 2 people drew the same regions for all three tasks. What were the sizes of the three regions? Given default instructions, where do people assign boundaries to a cognitive region they understand has no crisp boundary? We did not have any particular prior guess about this. In fact, 12 participants drew the 50% region as larger than the default, 12 drew it as smaller, and 12 drew them as the same size. That is, equal numbers of participants interpreted default instructions as weaker than 50% confidence, stronger than 50% confidence, or equal to 50% confidence. In contrast, no participant drew the 100% region as larger than the default; 32 drew it as smaller, and 4 drew it the same size. These results suggest that on average, people interpret a default

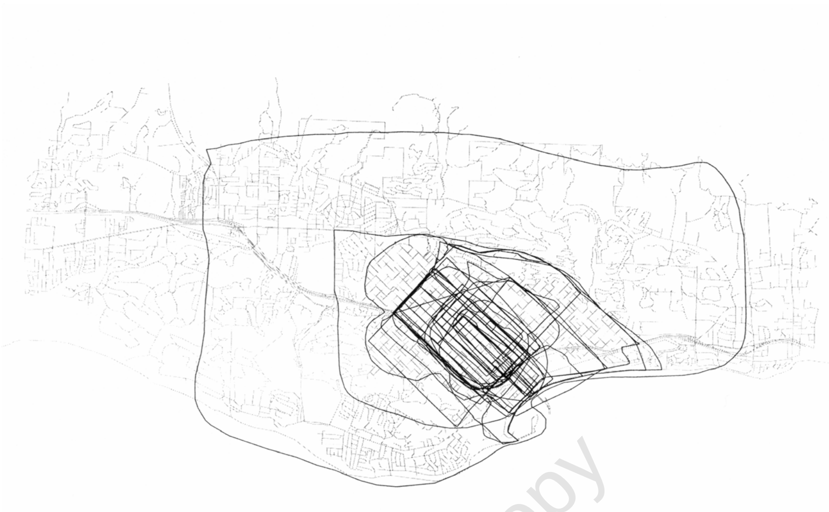

Figure 2. Raw-data polygons for each participant's default concept of downtown Santa Barbara.

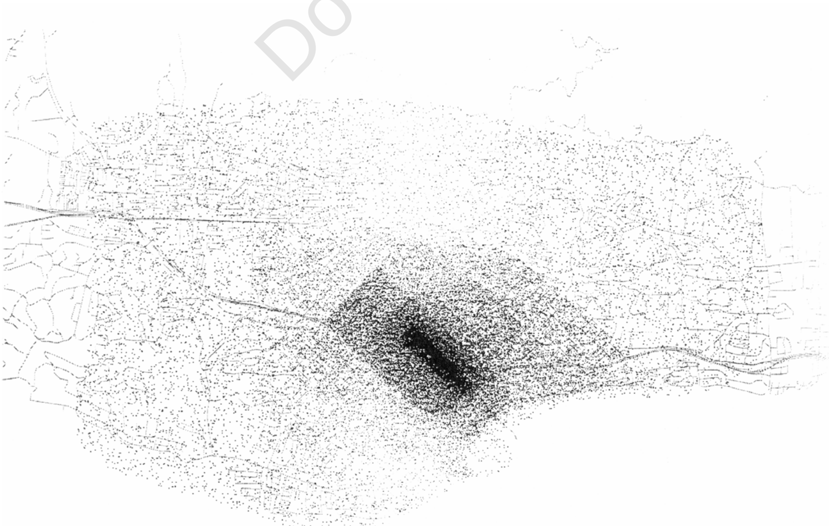

Figure 3. Default downtown region displayed with dot-density shading.

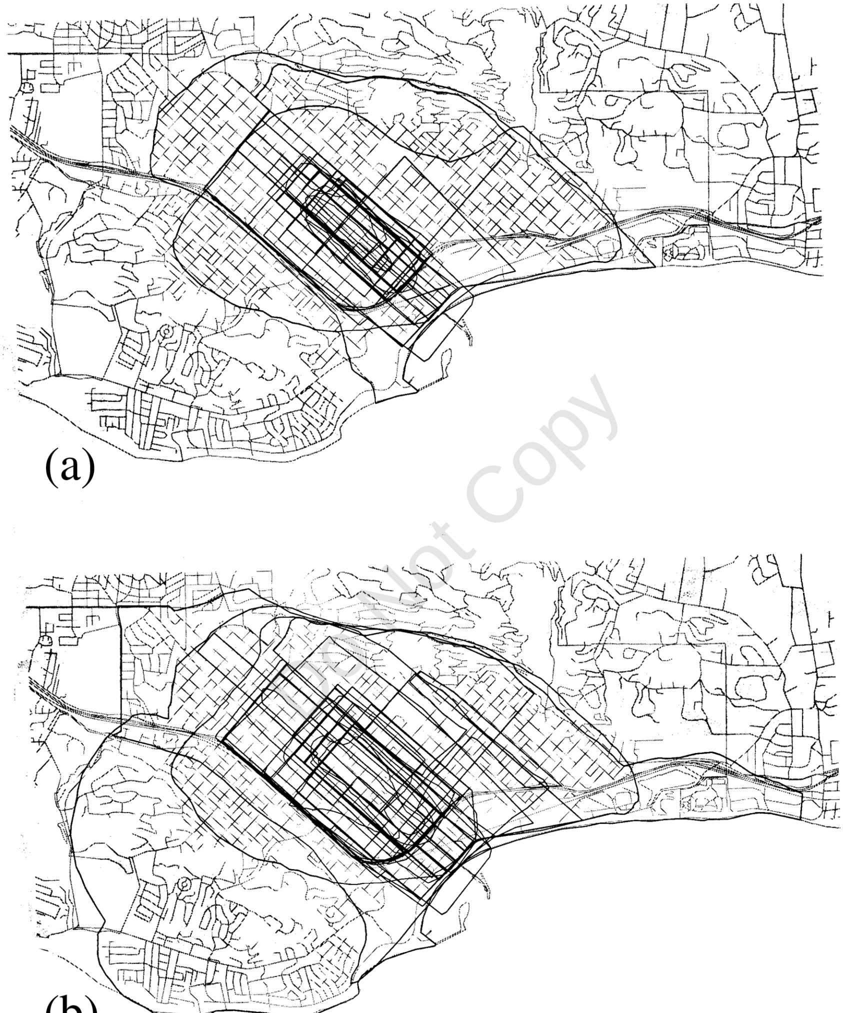

Figure 4. Raw-data polygons for each participant's a) 100% and b) 50% confidence downtowns.

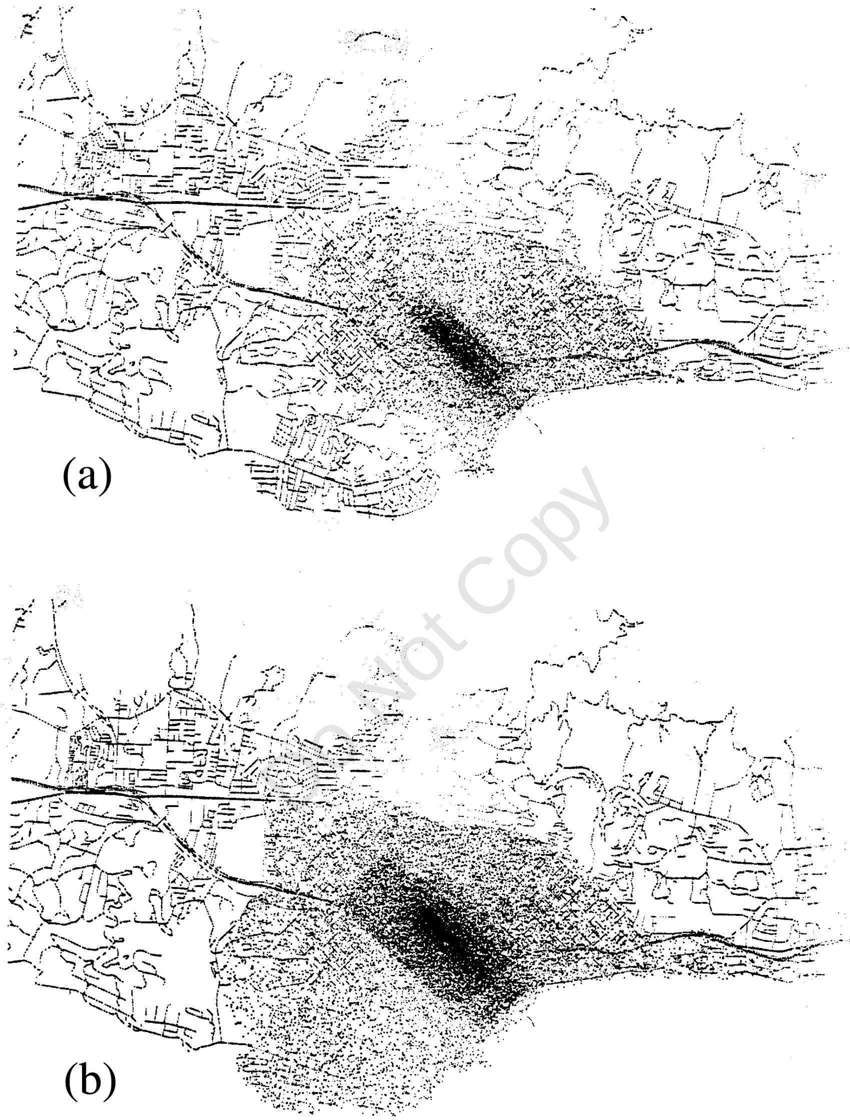

Figure 5. Confidence regions displayed with dot-density shading: a) 100% confidence.

request for *downtown* as calling for a boundary near the line of their 50% confidence. In general, smaller regions were wholly contained within larger regions, and no participant drew completely noncontiguous regions. But 9 participants did draw smaller or equal-sized regions that were only partially contained within the other region (typically 50% and/or 100% regions that only partially overlapped the default region).

In addition to the sizes and locations of the downtown regions, we examined their shapes. Twenty-three participants drew only convex polygons, with either rounded or rectangular corners. The remaining 13 participants drew at least one of their regions as concave, 8 drawing all of their regions as concave. Comments made by participants provided some insight into their bases for determining their regions, including their shapes. The presence of commercial enterprises was a common reason, consistent with a dictionary definition of downtown as central business district. But several people mentioned that downtown was the area in which people (presumably tourists) would walk. And one participant stated that the presence of City Hall and the courthouse defined downtown.

Participants readily understood our request for a core location in downtown. And the core was located in the most stringently defined downtown, the 100% confidence region, for all but 2 participants (2 others did not answer this question). As stated above, we tried to avoid suggesting to participants that the core had to be the spatial centroid of their regions. In fact, participants did not necessarily interpret the instructions to be a request for a spatial centroid. The core was located an average of nearly 300 meters from the centroid of the 100% region (to understand the scale of this, the average 100% region was only 600-800 m in width). It appears that participants' concepts of downtown were graded (whether fuzzy or probabilistic) but not symmetric around the core of downtown.

Summary and Conclusions

Many commentators have claimed that spatial information systems will benefit from applying knowledge of human conceptualizations of space and place to system design. There is little doubt that these conceptualizations are replete with categories, both spatial and nonspatial, that are vague, having imprecise and graded boundaries. Attempts to understand these conceptualizations must involve both theoretical and empirical work on the structures and processes of vague entities such as cognitive regions. In this paper, we have discussed some of the difficulties of representing vague regions in traditionally precise ways in digital systems. In addition to modeling and representing the structure of vague regions, it is important to develop methods for determining their content: To what do these terms refer? In this paper, we discussed empirical methods for answering this question in a specific context. Both a priori and interactive approaches were described. We reported a study in which participants drew lines around areas they believed constituted *downtown Santa Barbara*. Vagueness in the boundaries was elicited in two ways, by comparing variation in boundary locations across participants and by having participants draw different boundaries to indicate their varying confidence in region membership for different parts of the area. The results provide evidence that our method is a viable approach to externalizing people's representations of vague cognitive regions.

Interactive and a priori approaches to determining the referents of vague spatial queries are based on similar empirical methods. There are important differences between them. Interactive approaches are much more labor intensive for the user querying a database, but they allow determination of idiosyncratic meanings for any particular user. A priori approaches require prior data collection but do not place extra demand on the user at the time the query is placed. However, they require that there is sufficient agreement among different users that the stored vague concept applies generally, not just to a single user. In his study of vague spatial relations, Robinson (2000) concluded that there was little consistency among users, that it was "apparent from these results that one can expect little agreement among individual users on the exact definition of simple spatial relations" (p. 140). Our results suggest instead that people's conceptual understandings do reflect a fair amount of agreement, though clearly two people are unlikely to agree exactly on the meaning of a vague region like *downtown* (indeed, that is the frequentist notion of vagueness). Still, a consideration of the ultimate idiographic nature of conceptual structure may support the viability of an interactive approach over an a priori approach.

All approaches depend on the willingness of people to make discrete and precise judgments about region membership, even though they do not conceptualize boundaries as discrete and precise. Our study found that people were in fact quite willing and able to draw discrete boundaries around downtown, even though we also found that most people readily accepted the notion that downtown does not actually have a single precise boundary but a "band" of area of diminishing membership strength around a high confidence core (or possibly multiple discrete boundaries varying in degree of membership strength). Comments we recorded from participants indicated that they clearly believed downtown has a "core" location of greatest prototypicality, though not necessarily at the spatial centroid.

Of course, a person's concept of downtown is not a single context-free representation. The concept of "downtown" as expressed in thought and behavior (including language and mapping) is not due just to a static representation (or set of representations) stored in long-term memory. When a person uses knowledge about regions (or other conceptual entities), long-term memory representations are activated into working memory. The representations are thus subject to various processes of memory integration, transformation, etc., that take account of the particular context of the situation (including the purpose of querying about downtown, the form by which downtown is expressed, whether one is speaking to a local resident or a tourist, and so on). The precise referent of "downtown" may thus vary somewhat depending on contextual factors. The degree and nature of this variation is an empirical possibility; it should not be assumed necessarily to be great in magnitude. Questions about the effects of contextual factors make up critical research agenda for cognitive- and information-science communities.

An important example of such a contextual factor in the present case concerns the determination and presentation of appropriate base maps that include the entire area that might, even slightly, be considered to be part of a particular vague region. It is possible that using a base map could introduce some type of a map bias. For example, the area covered by the base map might influence the size of estimated regions (e.g., a smaller downtown might be elicited if the base map depicted less area). One could avoid such a bias by using a method not dependent on a base map. For example, one could stop participants at many different places around town and ask them to say whether they were downtown or not at that moment. From a research perspective, this would be an interesting technique to explore, though very labor intensive. However, its applicability to the situation of a user sitting at a terminal requesting spatial information is limited.

Natural language is inherently vague and imprecise; human communication often proceeds by a series of iterations when greater precision is needed. There is no single reason for the vagueness of conceptual structure—both epistemological and ontological vagueness are characteristic of human thought (though precision and certainty do occur as well). The user entering a map library without a precise definition of need enters into a dialog with an assistant in which both iterate towards an agreement. The assistant helps the user to refine the need and to identify the best solution in the form of an information object. So far, systems like digital libraries work fundamentally differently than this in their assumption that the user is able to approach the system with a precise need and that all of the objects available can be characterized precisely.

The tension between precise and vague specification extends far beyond the context of digital libraries to many other aspects of the interaction of society with information and the environment. Standard digital systems are inherently precise, requiring everything to be reduced to a binary alphabet. Traditional maps are also precise, forcing gradual transitions between regions to appear as sharp boundaries. Precision is relatively easy to achieve in a centralized authoritarian system where uniformity can be imposed. Thus an obvious solution to the *downtown Santa Barbara* problem posed in this paper would be to establish a standard that applies to all users of digital information systems, or even to society as a whole. But the precise agreement reached between the individual user and the map library assistant is a standard for two people only, and very different in its implications for the broader community. Thus the challenge for designers of digital libraries and other information systems is to achieve precision in a user's interaction with the system, without at the same time forcing that version of precision on all users.

Acknowledgments

The National Center for Geographic Information and Analysis (NCGIA), and the Alexandria Digital Library, are supported by the National Science Foundation. The National Imagery and Mapping Agency University Research Initiative also provided support for this work. Thanks to David Mark, the editors of this special issue, and anonymous reviewers of an earlier version for helpful comments.

References

Aitken, S. C., & Prosser, R. (1990). Residents' spatial knowledge of neighborhood continuity and form. *Geographical Analysis, 22,* 301–325.

Allen, G. L. (1997). From knowledge to words to wayfinding: Issues in the production and comprehension of route directions. In S. C. Hirtle & A. U. Frank (Eds.), *Spatial information theory: A theoretical basis for GIS*, COSIT'97, Vol. 1329 Lecture Notes in Computer Science (pp. 363–372). Berlin: Springer-Verlag.

Altman, D. (1994). Fuzzy set theoretic approaches for handling imprecision in spatial analysis. *International Journal of Geographical Information Systems, 8,* 271–289.

Blakemore, M. J. (1984). Generalization and error in spatial databases. *Cartographica, 21,* 131–139.

Brown, R. F. (1991). *Delimiting the perceived downtown: A perceptual approach*. Unpublished doctoral dissertation, The University of Pennsylvania, Philadelphia.

Burrough, P. A. (1996). Natural objects with indeterminate boundaries. In P. A. Burrough & A. U. Frank (Eds.), *Geographic objects with indeterminate boundaries* (pp. 3–28). London: Taylor & Francis.

Burrough, P. A., & Frank, A. U. (Eds.) (1996). *Geographic objects with indeterminate boundaries*. London: Taylor & Francis.

Cobb, M., Petry, F., & Robinson, V. E. (2000). Special issue on "Uncertainty in geographic information systems and spatial data." *Fuzzy Sets and Systems, 113*.

Cohn, A. G., & Gotts, N. M. (1996). The 'egg-yolk' representation of regions with indeterminate boundaries. In P. A. Burrough & A. U. Frank (Eds.), *Geographic objects with indeterminate boundaries* (pp. 171–188). London: Taylor & Francis.

De Bruin, S. (2000). Querying probabilistic land cover data using fuzzy set theory. *International Journal of Geographical Information Science, 14,* 359–372.

Frank, A. U. (1996). Qualitative spatial reasoning: Cardinal directions as an example. *International Journal of Geographical Information Systems, 10,* 269–290.

Goodchild, M. F., & Proctor, J. (1997). Scale in a digital geographic world. *Geographical & Environmental Modelling, 1,* 5–23.

Gray, M. V. (1997). Classification as an impediment to the reliable and valid use of spatial information: A disaggregate approach. In S. C. Hirtle & A. U. Frank (Eds.), *Spatial information theory: A theoretical basis for GIS*,, Vol. 1329 Lecture Notes in Computer Science (pp. 137–149). Berlin: Springer-Verlag.

Kuhn, W. (2001). Ontologies in support of activities in geographical space. *International Journal of Geographical Information Science, 15,* 613–631.

Lakoff, G. (1987). *Women, fire, and dangerous things.* Chicago, IL: University of Chicago Press.

Lavin, S. (1986). Mapping continuous geographical distributions using dot-density shading. *The American Cartographer, 13,* 140–150.

Mark, D. M. (1993). Toward a theoretical framework for geographic entity types. In A. U. Frank & I. Campari (Eds.), *Spatial information theory: A theoretical basis for GIS*, Vol. 716 Lecture Notes in Computer Science (pp. 270–283). Berlin: Springer-Verlag.

Mark, D. M., & Csillag, F. (1989). The nature of boundaries on "area-class" maps. *Cartographica, 26,* 65–78.

Mark, D. M., & Frank, A. U. (1989). Concepts of space and spatial language. *Proceedings, Ninth International Symposium on Computer-Assisted Cartography* (pp. 538–556). Baltimore, MD.

Montello, D. R. (2003). Regions in geography: Process and content. In M. Duckham, M. F. Goodchild, & M. F. Worboys (Eds.), *Foundations of geographic information science* (pp. 173–189). London: Taylor & Francis.

Montello, D. R., & Frank, A. U. (1996). Modeling directional knowledge and reasoning in environmental spaces: Testing qualitative metrics. In J. Portugali (Ed.), *The Construction of Cognitive Maps* (pp. 321–344). Dordrecht: Kluwer Academic.

Papadias, D., Karacapilidis, N., & Arkoumanis, D. (1999). Processing fuzzy spatial queries: A configuration similarity approach. *International Journal of Geographical Information Science, 13,* 93–118.

Pickles, J. (Ed.) (1995). *Ground truth: The social implications of geographic information systems*. New York, NY: Guilford.

Retz-Schmidt, G. (1988). Various views on spatial prepositions. *AI Magazine, Summer,* 95–105.

Robinson, V. B. (1990). Interactive machine acquisition of a fuzzy spatial relation. *Computers and Geosciences, 16,* 857–872.

Robinson, V. B. (2000). Individual and multipersonal fuzzy spatial relations acquired using human-machine interaction. *Fuzzy Sets and Systems, 113,* 133–145.

Rosch, E. H., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605.

Smith, B., & Varzi, A. C. (1997). Fiat and bona fide boundaries: Towards an ontology of spatially extended objects. In S. C. Hirtle, & A. U. Frank (Eds.), *Spatial information theory: A theoretical basis for GIS*, Vol. 1329 Lecture Notes in Computer Science (pp. 103–119). Berlin: Springer-Verlag.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Smith, T. R., Andresen, D., Carver, L., Dolin, R., Fischer, C., Frew, J., et al. (1996). A digital library for geographically referenced materials. *Computer, 29,* 14.

Streeter, L. A., Vitello, D., & Wonsiewicz, S. A. (1985). How to tell people where to go: Comparing navigational aids. *International Journal of Man/Machine Studies, 22,* 549–562.

Wahlster, W. (1989). One word says more than 1000 pictures—On the automatic verbalization of the results of image sequence analysis systems. *Computers and Artificial Intelligence, 8,* 479–492.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General, 115,* 348–365.

Wang, F. (1994). Towards a natural language user interface: An approach of fuzzy query. *International Journal of Geographical Information Systems, 8,* 143–162.

Wang, F., & Hall, G. B. (1996). Fuzzy representation of geographical boundaries in GIS. *International Journal of Geographical Information Systems, 10,* 573–590.

Zadeh, L. A.(1965). Fuzzy sets. *Information and Control, 8,* 338–353.