Ontology with Human Subjects Testing:

An Empirical Investigation of Geographic Categories

Barry Smith[1] and David Mark[2]

Abstract

The paper presents a framework for the formulation and testing of ontological theories embodied in human cognition, concentrating primarily on the domain of geographic categories. Evidence for and against alternative theories of cognitive categories, for example on the part of E. Rosch and her associates, has been hitherto based primarily on studies of categorization of entities of table-top space (*pets*, *tools*, *fruits*). We hypothesize that the structure of our categories does not remain constant as we move from categories of objects at manipulable scales to geographic categories such as *nation, mountain, river.* More precisely: Geographic objects are not merely *located* in space, they are tied intrinsically to space in such a way that they inherit from space many of its structural (mereological, topological, geometrical) properties. Categorization in the geographic world is often size- or scale-dependent (consider: *pond*, *lake*, *sea*, *ocean*), and to a much greater extent than in the world of table-top space, the realization that a thing or type of thing exists at all in the geographic world may have individual or cultural variability. Geographic objects are in very many cases the products of delineation within a continuum, and the boundaries of such objects are themselves highly salient phenomena for purposes of categorization. A battery of experiments is described to test these hypotheses and to serve as a basis for more detailed ontological theorizing.

1.   Introduction

Ontology, since Aristotle, has been conceived as a sort of highly general physics, a science of the types of entities in reality, of the objects, properties, categories and relations which make up the world. At the same time ontology has been for some two thousand years a speculative enterprise. It has rested methodologically on introspection and on the construction and analysis of elaborate world-

[1]Department of Philosophy, NCGIA and Center for Cognitive Science, University at Buffalo. Email: phismith@acsu.buffalo.edu.

[2]Department of Geography, NCGIA and Center for Cognitive Science, University at Buffalo. Email: dmark@geog.buffalo.edu.

models and of abstract formal-ontological theories. In the work of Quine and others this ontological theorizing in abstract fashion about the world was supplemented by the study, based on the use of logical methods, of the ontological commitments or presuppositions embodied in scientific theories. In recent years both types of ontological study have found application in the world of information systems, for example in the construction of frameworks for knowledge representation and in database design and translation. As ontology is in this way drawn closer to the domain of real-world applications, the question arises as to whether it is possible to use empirical methods in studying ontological theories. More specifically: can we use empirical methods to test the ontological theories embodied in human cognition? In what follows we set forth the outlines of a framework for the formulation and testing of such theories as they relate to the specific domain of geographic objects and categories.

Objects, properties, categories and relations are what they are, independently of how people think of them. Some objects, properties, categories and relations, however, are the *products* of human cognition. This holds not least in the geographic realm, where many of the entities with which we have to deal may be conceived by analogy with shadows cast on the surface of the earth by human practices of specific sorts. In relation to such entities empirical testing makes reasonable sense. We describe a testing methodology in which the more traditional methods of ontology will guide the formulation of questions to be tested and the construction of the framework in which the results of testing shall be expressed.

2.   Theories of Conceptual Organization

We begin with the general topic of human cognitive categories such as *rabbit, electron, island*. Such categories exist in two forms: on the one hand as concepts on the side of human subjects; on the other hand as kinds on the side of reality. On the classical view, dating back to Aristotle, each concept or kind is associated with certain defining attributes or properties which suffice to determine exactly which objects fall within the relevant extension. On more recent views, categorial kinds are to be understood by analogy with a mathematical set. All objects within the extension set are equally representative instances of the category, and for each object or event it is fully determinate whether or not it falls under a given category.

Geographers, like other scientists, have typically accepted this model of categories as sets in the mathematical sense, and the model is presupposed for example in work on cartographic data standards (see Mark 1993, 1993a). As an account of the categories used by ordinary humans in everyday situations, however, the model has obvious defects. First, and most obviously, not every

set in the mathematical sense is a class in the sense of kind or category. Hence we need to go beyond set theory in order to fill this gap. But further, as has been shown by Rosch (1973, 1978) and others (see for example Keil 1979, Estes 1994), for most such categories, and for most people, some members are better examples of the class than are others; furthermore, there is a great degree of agreement among human subjects as to what constitute good and bad examples. Human cognitive categories often possess a radial structure, having prototypes or more central or typical members surrounded by a penumbra of less central or less typical instances. Sparrows and crows are more, ostriches and flamingoes less typical instances of the category *bird*.

Rosch raised the following question: Why do children learn so readily category-terms like *duck*, *zebra*, *clock*, *fork* while they experience difficulties learning terms like *mammal* or *utensil*? The former list of terms belongs to what she calls the 'basic level' of cognitive classification, the level on which categories most easily learned in given domains of discourse are to be found. This basic level is a compromise between two opposing goals, that of informativeness, and that of minimizing categories based on irrelevant distinctions. The basic level (*chair*, *apple*) thus falls between the superordinate level (*furniture*, *fruit*), which is in general insufficiently informative, and the subordinate level (*lounge chair*, *golden delicious*), which adds too little informativeness for its additional cognitive cost. Measures of our perception of stimuli, of our responses to stimuli, and of our communication, all converge on the same basic level.

3.   The Special Case of Geographic Categories

Psychologists and other cognitive scientists have developed a range of alternative theories about the nature of categories, of which Rosch's prototype theory is one important example. Evidence for and against these various theories is, however, based almost entirely on empirical studies of categorization of entities of table-top space, such as small pets, tools, and other manipulable artifacts, or of abstract entities (properties) such as colors and diseases. A question that has not been raised is whether the structure of our categories is constant as we move beyond examples derived from objects at table-top or manipulable scales and turn to examples derived from the sphere of geographic objects. Are there peculiarities of geographic categories such as *nation, mountain, river,* that set them apart from categories of other sorts? Such peculiarities might include:

(1) Geographic objects are not merely *located* in space, they are tied intrinsically to space in such a way that they inherit from space many of its structural (mereological, topological, geometrical) properties. For entities on the sub-geographic scale, the 'what' and the 'where' are almost always independent. In the geographic world, in contrast, the 'what' and the 'where' seem to be much more closely intertwined.

(2) Categorization in the geographic world is often size- or scale-dependent (consider: *pond*,

lake, *sea*, *ocean*).

(3) To a much greater extent than in the world of table-top space, the realization that a thing or type of thing exists at all in the geographic world may have individual or cultural variability.

(4) Geographic objects are in very many cases the products of delineation within a continuum, a continuum through or within which other types of objects, including human agents, live and move.

(5) The boundaries of the objects with which we have to deal in the geographic world are themselves salient phenomena for purposes of categorization. These boundaries may be crisp or graded, and they may be subject to dispute. Moreover, the identification of what a thing is may influence the location and structure of the boundary. For example, if something is a marsh, its boundary may be further up the slope than if the same thing is considered to be a lake.

4.   **Towards an Ontological Theory of Geographic Objects**

In what follows we shall use 'entity' and 'object' synonymously as ontological terms of art comprehending things, relations, boundaries, events, processes, qualities, quantities of all sorts. More specifically, in the context of geographic ontology, 'object' and 'entity' shall comprehend regions, boundaries, parcels of land and water-bodies, roads, buildings, bridges, and so on, as well as the parts and aggregates of all of these.

Geographic objects are spatial objects on or near the surface of the earth. Furthermore, they are objects of a certain minimal scale (roughly: of a scale such that they cannot be surveyed unaided within a single perceptual act). Geographic objects are typically complex, and they will standardly have parts. An adequate ontology of geographic objects must therefore contain a theory of part and whole, or *mereology* (Simons 1987).

Geographic objects do not merely have constituent object-parts, they also have boundaries, which contribute as much to their ontological make-up as do the constituents that they comprehend in their interiors. Geographic objects are prototypically connected or contiguous, but they are sometimes scattered or separated. They are sometimes closed (e.g., lakes), and sometimes open (e.g., bays). The above concepts of contiguity and closure are topological notions, and thus an adequate ontology of geographic objects must contain also a *topology*, a theory of boundaries and interiors, of connectedness and separation. The latter must be integrated with a mereological theory of parts and wholes to form a 'mereotopology' (Smith 1996, Smith and Varzi 1997), a theory able to do justice to the fact that spatial regions form a relational system, comprising also containment relations, separation relations, relations of adjacency and overlap, and so on (Egenhofer and Herring 1991; Mark and Egenhofer 1994a; Cohn and Gotts 1994).

An object is 'closed' in the mereotopological sense, if it includes its outer boundary as part; it is 'open' if this outer boundary is included rather in its complement. Ordinary material objects are in unproblematic fashion the owners of their surfaces. Where a complement meets an object of this sort, the object will be closed and the complement open (Asher and Vieu 1995). Regarding geographic objects, however, matters are not so simple. Consider the mouth of a bay, where the hole meets the open sea. Here a choice as to where we place the boundary would seem arbitrary, and a parallel situation is encountered vis-à-vis the borders separating hills and valleys. This arbitrariness seems to be an especially common feature of the geographic world, and we hypothesize that it will imply important features of geographic objects in general and of their boundaries in particular.

Geographic categories track not only mereology and topology but also *qualitative geometry* (the theory of concavity, convexity, of shortness and longness, the theory of being roughly round or roughly dumbbell shaped; see Cohn et al. 1997, 1997a; Smith and Varzi, in press). A theory of geographic categories must include, too, a *theory of dimension*, since it is a highly salient feature of such objects that they may be zero-, one-, two- or three-dimensional. Consider the North Pole, the Arctic Circle, Norway (a two-dimensional object with a curvature in three-dimensional space), or the North Sea. It is an important feature of many geographic terms that they may allow a switching from one sort of dimensional representation to another. Thus 'North Sea' may refer either to the three-dimensional body of water, or to the two-dimensional surface. Such shifting of reference implies also an analogous shifting on the level of conceptual categories. 'Bay' or 'sound' may refer to the surrounding land, or to the indentation in the shoreline, or to a part of the shoreline, as well as the sheet or body of water. There are correspondingly different meanings of 'in' (and of other spatial prepositions) according to what the relevant dimension in a given context might be: the island is 'in' the lake means that it protrudes from the surface of the lake; the submarine is 'in' the lake means that it is completely submerged within the corresponding three-dimensional volume.

An ontological theory of geographic objects must include further a theory of location, or more precisely a theory of the relation of *being located at* which holds between things on the one hand (roads, forests, wetlands), and the regions in or at which they are located on the other (Casati and Varzi 1996).

5.   Geographic Categories, Types of Predication, and Scale Continua

Ontologists since Aristotle have distinguished between two sorts of predications: *categorial predications* as we are here using this term (called by Aristotelians 'predications in the category of substance'), for example: is a man, is a fish, is a lake, etc.; and *accidental predications* (or 'predications in the category of accident'), for example: is red, is colored, is big, is hungry. The former tell us under what category an object falls. They tell us *what an object is*. The latter tell us

how an object is, *per accidens*, at a given moment, *what state the object is in*, or *what process it is undergoing*; thus they pertain to ways in which instances of the relevant categories change from occasion to occasion.

For objects of table-top space, now, predications of location and size are almost always accidental: practically all such objects can move, and animate objects (which are among the most salient) change size in regular ways over time. Terms for basic-level categories at table-top scales will thus not code for location or size. They will not code, either, for position (table-top categories do not change category, for example, when they are upside down). Within the realm of geographic objects, however, matters are quite different. Here, at least within the time-scales relevant to the development of human cognitive capacities, almost all objects do not move or grow. Size, shape and position may thus be matters for categorical predication. Good candidate basic-level terms will therefore often form series, as illustrated by the case *pond – lake – sea – ocean, bay – cove*, *mountain – hill – hillock*, of a sort which seem to be common for geographic categories and rare elsewhere.

6.   Problems with Geographical Extensions of Theories of Categorization Based on the Phenomena of Table-Top Space

Our cognitive acts are directed towards spatial objects in the world. But these acts themselves exist in the spatial domain in virtue of the fact that they are tied to our bodies, so that some of our spatial concepts, like *here* or *there*, are egocentric. In contrast to other families of categories, therefore, conceptual categories in the spatial realm relate to their objects in manifold fashion: i) through abstract models or representations of space in our minds, as when we think, abstractly, about whether the Bay of Biscay is to the North or to the South of Long Island; ii) through a concrete being-in-space, as when we use indexical spatial concepts like *yonder*, *to the right*, *down east*, etc.; and iii) through different sorts of combinations of these two. Matters are complicated still further by the fact that, because we are in space, are surrounded by space, and are not able to manipulate space itself, our cognitive representations of space may be underdefined or erroneous. Objects of geographic categorization are too large to be taken in within a single act of perception, and thus *a fortiori* they are too large to serve easily as targets of comparison. Some theory, and much additional contextual knowledge will be required for categorization purposes, and for this reason, too, geographic categories may be expected to show marked individual or culture-related differences. There is, incidentally, an analogous problem in regard to the use of normal, ostensive means in referring to geographic objects. As Bennett writes:

> there is no practicable way of giving a hurricane the name 'Gloria' unless you say something like 'Gloria is the hurricane that ...'. You might stand in the middle of the hurricane, wave your arms, and shout '*This* is Gloria', but the rest of us don't know how far your 'this' is

meant to reach, and so we don't know what you are calling 'Gloria' (1988, p. 3).

The table-top examples that have traditionally been treated in the literature on categorization differ from geographic examples also in other respects. First, they almost always involve discrete, movable items, items which can be observed from all sides, items which do not change category when inverted. And while research on categorization by cognitive scientists does indeed indicate that humans tend quite generally to discretize, even where it is essentially continuous phenomena which are at issue, an adequate ontology of geographic kinds should embrace not only categories of discreta but also categories that arise in the realm of continuous phenomena, above all categories of large-scale continuous phenomena within which discrete objects are contained or located and through or along which they can move. (See Brogaard, Peuquet and Smith 1999.)

The theoretical concentration on independently movable and manipulable, table-top examples tend further to reinforce a view according to which nature can be cut at its joints—that is, a view to the effect that there is a true, God-given structure, which science attempts to make precise. As we have seen, however, geographic categorization involves a degree of human-contributed arbitrariness on a number of different levels, and it is in general marked by differences in the ways different languages and cultures structure or slice their worlds. It is precisely because, as we hypothesize, many geographical kinds result from a more-or-less arbitrary drawing of boundaries in a continuum that the category boundaries will likely differ from culture to culture, in ways that can, under some conditions, lead to conflict between one group or culture and another. (See Smith 1997.)

Finally, the most often studied table-top examples form a family of separate categorial systems possessing simple genus-species tree structures organized in terms of greater and lesser generalities, each tree having little to do with the other trees. Thus, for example, the category *bird* and its sub-categories will have little interaction with the category-families *utensil* or *item of clothing*. We hypothesize that geographic categories, in contrast, because they relate to objects intrinsically interrelated together within a single domain (called *space*), form categorial systems that interact more intimately to form a single structure. Thus many geographic categories form mutually interdependent pairs (*hill/valley, land/water, bay/promontory*) in a way which is rare among standard categories of objects in a space of table-top extent.

7.   **The Realm of Fiats**

Geographic objects will often be identified by defining the locations of their boundaries. We have distinguished two kinds of boundaries: bona fide boundaries and fiat boundaries (Smith 1994, 1995, Smith and Varzi 1997). *Bona fide boundaries* are those that correspond to genuine discontinuities in the world; *fiat boundaries* are projected into geographic space at locations wholly or partly

independent of such discontinuities. The surfaces of extended objects such as lakes or islands are boundaries of the bona fide sort. Roads and water courses can also readily be considered to be bona fide boundaries. In contrast, most state and provincial borders, as well as many county and property lines and the borders of census tracts and of postal and electoral districts, provide examples of outer boundaries of the fiat sort, especially in those cases where, as in the case of Colorado or Utah, they lie skew to any pre-existing qualitative differentiations or spatial discontinuities (coastlines, rivers) in the underlying territory. Boundaries of areas of some given soil type, of wetlands, or of bays or mountains are also at least partly of the fiat type, although they may result from cognitive rather than legal-administrative processes. Bona fide boundaries exist in all domains of reality, from the microphysical to the cosmological. Fiat boundaries are found and are relevant to categorization almost exclusively in the realm of geographic entities and in cognate realms of law, politics and political administration.

Moreover once fiat boundaries have been recognized, it becomes clear that the opposition between bona fide and fiat boundaries implies a parallel opposition also in relation to boundaries but in relation to the objects that they bound. Examples of bona fide geographic objects are the planet Earth, Vancouver Island, and the Dead Sea. Examples of fiat geographic objects include King County, the State of Wyoming, and the Tropic of Capricorn.

8.   Types of Fiat Boundaries

Fiat boundaries in the geographic realm come into being in virtue of different sorts of demarcations effected cognitively by human beings. There are fiat objects (deserts, valleys, etc.) that are delineated not by crisp outer boundaries, but rather by boundary-like regions that are to some degree vague or indeterminate. Such vagueness is a conceptual matter: if you point to an irregularly shaped protuberance in the sand and say 'dune', then the correlate of your expression is a fiat object whose constituent unitary parts are comprehended (articulated) through the concept *dune*. The vagueness of the concept itself is responsible for the vagueness with which the referent of your expression is picked out.

Many obvious examples of fiat objects involve cases where proper parts are delineated or carved out (by fiat) within the interiors of larger bona fide wholes. Consider the way in which multiple nations and states may be carved out within a single continental land mass. But there are also fiat wholes—for example New England, Benelux, the European Union—that are created by the conjoining of multiple parts into a single composite whole. It seems that we can reasonably assume of bona fide objects (a person, a rock, the planet Earth) that they are connected in the topological sense (the solar system may be an exception to this rule); fiat objects, however, may quite generally be scattered: they may, like Polynesia, be such as to include non-connected bona fide objects within

larger fiat wholes.

9. Fiats in the Realm of Categories

The concept of fiat boundary was introduced as a means of doing justice to the fact that we divide up the spatial reality out there in more or less arbitrary fashion into sub-regions. But there is an element of arbitrariness or fiat also in the domain of our categories themselves: we can partition the family of spatial categories in more or less arbitrary ways into sub-categories. Thus, for example, erms like *strait* and *river* represent arbitrary partitions of the world of water bodies. The English language might have evolved with just one term, or three terms, comprehending the range of phenomena stretching between *strait* and *river* or, in French, between *détroit* and *fleuve*. For while the Straits of Gibraltar are certainly not a river, and the Mississippi River is certainly not a strait, there are cases—such as the Detroit and Niagara Rivers and the Bosporus—that exist on the borderline between the categories. All are flat, narrow passages that ships can sail through between two larger water bodies (lakes, seas), and all have net flow through them, due to runoff, etc.

Imagine the instances of a concept arranged in a quasi-spatial way, as happens for example in familiar accounts of color- or tone-space. Suppose that each concept is associated with some extended region within this quasi-space in which its instances are contained, and suppose further that this is done in such a way that the prototypes, the most typical instances, are located in the center of the relevant region, the less typical instances being located at distances from this center in proportion to their degree of non-typicality. Boundary cases can now be defined as those cases that are so untypical that even the slightest further deviation from the norm would imply that they are no longer instances of the given category at all.

In this way counterparts of the familiar topological notions of boundary, interior, contact, separation, and continuity can be defined for the realm of conceptual categories, and the notion of similarity as a relation between instances can be understood as a topological notion (Mostowski 1983, Petitot 1995). In the realm of colors, for example, "*a* is similar to *b*" might be taken to mean that the colors of *a* and *b* lie so close together in color-space that they cannot be discriminated with the naked eye. In some cases, there is a continuous transition from one concept to its neighbors in concept-space, as for example in the transition from peninsula to promontory or from lake to marsh to wetland. In other cases, categories are separated by gaps (by regions of concept-space that have no instances). This is so regarding the transition from, say, lake to reservoir.

10. Water in Geographic Space

Water is an especially distinctive substance that is critical to life. Water occurs at scales from individual water molecules in water vapor to the great body of water making up the world's oceans. Whereas water bodies and water courses make up some of the most distinctive geographic categories, there nevertheless is a gradation from lakes to wetlands to dry land, and boundaries of individual water bodies may grade into fringing marshes and thence to terrestrial habitats.

Consider, for example, the case of lakes. Is a lake a three-dimensional body of water in geographic space, or a two-dimensional sheet of water, or is it a depression in the Earth's surface, (possibly) filled with water? Dry lakes exist, but are they lakes when they are dry, or merely places where lakes were, and might be again? One place to start is with definitions contained in geographic or cartographic data standards, or in dictionaries. These definitions represent the consensus among experts as to the meanings of terms that are used to refer to geographic categories. They thus provide preliminary evidence as to the nature of the categories themselves.

The U.S. Spatial Data Transfer Standard (SDTS) defines *lake* as "any standard body of inland water." (See Mark 1993, 1993a) The U.S. National Imagery and Mapping Agency (NIMA) defines *lake* as the water that composes it, but in the following way: "water contained within a predominantly natural shoreline that exhibits no appreciable current." A *lake*, then, on this ontology is a *body of water* of a certain sort. But matters are not so simple, as is made clear by the following reflections from the discussion of the status of lakes in the more general treatise on the ontology of liquids by Patrick Hayes (1985a):

> Consider now a lake. This is a contained-space defined by geographical constraints. Lake Leman, for example, is the space contained between the Jura Mountains, Lausanne, the Dent d'Oche, Thonon, and the Rochers de Naye, below the 400 meter contour (more or less). Its container is the surface of the earth under it, i.e. the *lake bed*. I think the only way to describe lakes, rivers and ponds in the present framework is to say that they are contained-spaces which are full of water: that is, the space ends at the surface of the water. To be *in* the lake is then, reasonably, to be immersed in water, while to be *on* the lake is to be immediately above the water and supported by the lake (cf. on the table), which seems reasonable. Thus a lake is full by definition.

A lake is never half full, if Hayes is right. Rather, if it contains only half of its usual volume of water; then its level is low. A reservoir behind a dam, in contrast, can be half full, or empty. This is a matter of the ontology of lakes. Hayes contrasts his view with the ontology of water bodies according to which ponds, lakes, seas, etc. are all pieces-of-water under a different name. On Hayes' view
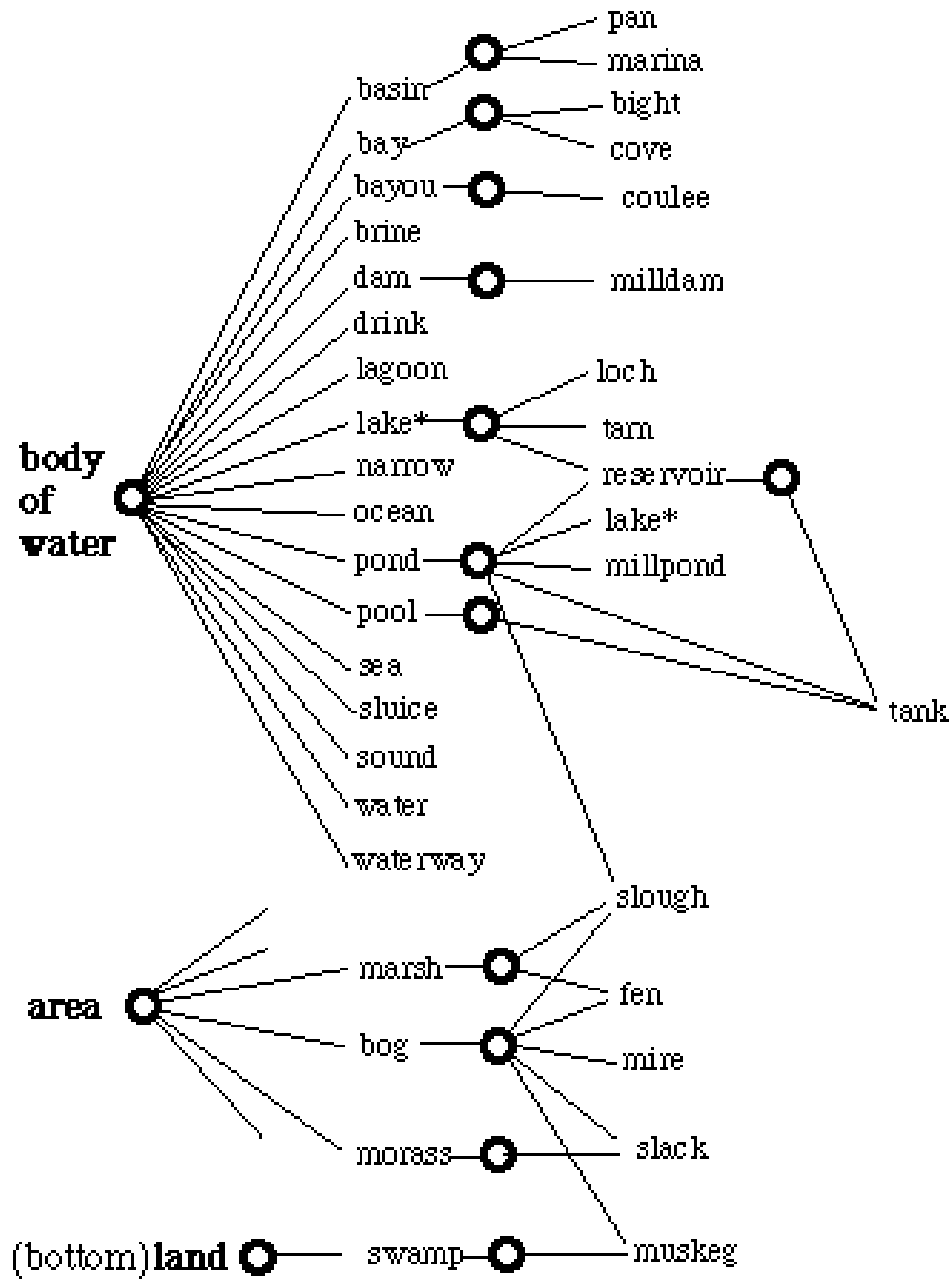
> Lake Leman is a fixed object in geographical space whereas in the pieces-of-water ontology, it would be constantly changing, since the Rhone flows in one end and out the other; it would be a phenomenon, not an object.

Is the Hayes ontology more or less adequate as a specification of the relevant portions of our categorial scheme than the pieces-of-water ontology? This is a question which admits of empirical testing.

11.   Dimensions for Categories of Geographic Water Bodies

Apart from size, shape and location, there is a range of further salient dimensions involved in our categorization of geographic objects. To gain a first assay of what these dimensions might be, we carried out a pilot study of those geographic categories which fall within the general class of water bodies. A computer-aided search of an electronic version of the *American Heritage Dictionary* found 121 definitions that include both the word "body" and the word "water," but only 73 terms contain both "body" and "water" in the same noun phrase. Furthermore, many of these referred to parts or denizens of water bodies, leaving 18 terms whose definition stated that they were "a body of water" with certain characteristics or restrictions: *basin*, *bay*, *bayou*, *brine*, *dam*, *drink*, *harbor*, *lagoon*, *lake*, *narrow*, *ocean*, *pond*, *pool*, *sea*, *sluice*, *sound*, *water*, and *waterway*. Figure 1 shows these terms linked to their superordinate category "body of water," and other terms in that dictionary that have these kinds of water bodies as superordinates.

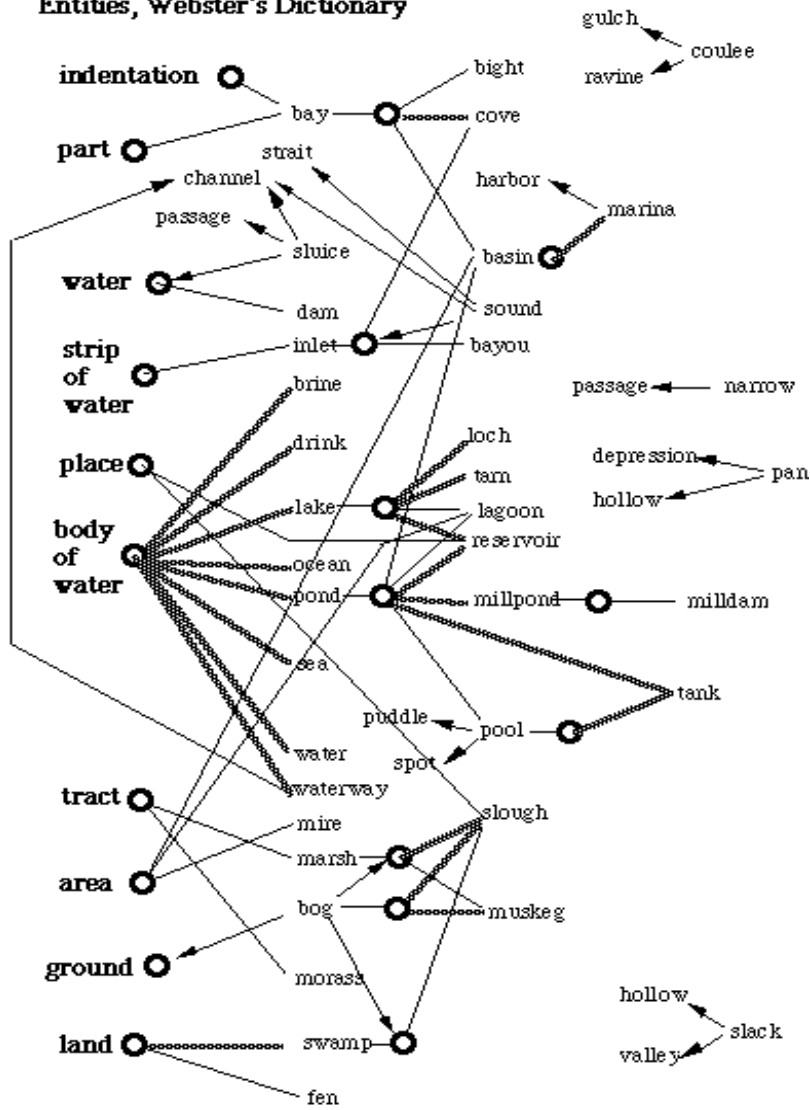# Ontology of Water Bodies and Related Entities, based on Definitions in the American Heritage Dictionary

basin — pan
basin — marina
bay — bight
bay — cove
bayou — coulee
brine
dam — milldam
drink
lagoon
lake* — loch
lake* — tam
narrow
ocean — reservoir — tank
pond — lake*
pond — millpond
pool
sea
sluice
sound
water
waterway

body of water

area — slough
area — marsh — fen
area — bog — mire
area — morass — slack
area

(bottom)land — swamp — muskeg

* = term appears twice

<u>Figure 1</u>

12

When the definitions of the same terms were examined in *Webster's Dictionary of the American Language*, a different semantic network emerged (Figure 2). The thicker gray links in Figure 2 indicate those subclass-superclass relations that were the same in both dictionaries.

Ontology of Water Bodies and Related Entities, Webster's Dictionary

gulch

coulee

ravine

indentation ○

bight

bay ○····· cove

part ○

strait

channel

harbor

marina

passage

sluice

basin ○

water ○

dam

sound

strip
of
water ○

inlet

bayou

brine

passage ← narrow

drink

loch

depression

place ○

tarn

hollow

pan

lake ○

lagoon

body
of
water ○

reservoir

ocean

pond ○

millpond ○

milldam

sea

tank

puddle

pool ○

water

spot

tract ○

waterway

mire

slough

marsh ○

area ○

bog ○

muskeg

ground ○

morass

hollow

slack

land ○···············

swamp ○

valley

fen

* = term appears twice

Figure 2

14

Looking at the definitions of the above terms, it is possible to identify some individual descriptive dimensions that the compilers of these dictionaries felt were important in distinguishing various kinds of water bodies. These dimensions are size, shape, flow, other water properties, spatial relations, use, purpose. These will be useful dimensions for interpreting distinctions made among kinds of water bodies and water courses in the empirical investigation of geographical categories.

12.   Previous Empirical Research on Geographic Categories

Although there have been many deductive works addressing classification of geographic objects and phenomena, including dictionaries of geographic terms and cartographic data standards, there have been very few empirical studies of geographic categories that have involved testing with human subjects. The earliest such study that we are aware of involved a small number of categories from Battig and Montague's (1968) study of category norms. Two other such studies were conducted more recently: Tversky and Hemenway's (1983) research on indoor and outdoor scenes, and Lloyd *et al*.'s (1996) investigation of basic-level geographic categories. These three studies are reviewed in this section, with emphasis on methods and results.

12.1   Battig and Montague's Research on Category Norms

Norms for a category are instances of that category that are most commonly given to exemplify the category itself; they may be exemplars or prototypes of the category, although this is not necessarily the case. Battig and Montague (1968) used an elicitation-of-examples procedure to determine norms for examples of 56 categories. Of the categories that they tested, a few were geographic in nature. More than 400 undergraduate subjects from Maryland and Illinois were given category titles, and asked to write down in 30 seconds as many "items included in that category as you can, in whatever order they happen to occur to you." The subjects went through all 56 categories in this manner. The researchers tabulated all terms listed, and counted how many times each term was given under each category, and how often it was the first term mentioned. They also reported correlations between the rankings by Illinois students and the rankings by Maryland students. Cross-site correlations were generally high, indicating high stability across the speech communities tested. The lowest three correlations were for a city (0.689), a state (0.297), and a college or university (0.097), indicating that variation in examples for categories involving geographic instances rather than categories was itself varying geographically.

Of the categories tested by Battig and Montague, one was "a Natural Earth Formation." A total of 34 different "earth formations" were listed by at least 10 of the subjects. The ten most frequently-listed terms, with their frequencies among 442 subjects, were: *mountain* (401), *hill* (227), *valley*

(227), *river* (147), *rock* (105), *lake* (98), *canyon* (81), *cliff* (77), *ocean* (77), and *cave* (69). Despite the fact that the category was not prefixed by "a kind of" or "a type of," only one particular named feature was listed: the Grand Canyon was mentioned 14 times. All other terms given 5 or more times were names of categories, and all but 5 were at a geographic scale. Nothing movable was on the list, except glacier (very slow moving) and iceberg.

12.2   Tversky and Hemenway's Research on Indoor and Outdoor Scenes

Tversky and Hemenway (1983) applied Rosch's research methods to objects of geographic scale – in their paper called '(outdoor) environmental scenes'. Their goal was to provide a taxonomy of kinds of environmental scenes and to identify a basic level of scene categorization, the level not only most commonly used, but also 'apparently most useful in other domains of knowledge concerned with environments, for example, architecture and geography'. Forty-seven students at Stanford University served as subjects in two sets of experiments. In the first, subjects were presented with slides depicting common and familiar indoor and outdoor scenes and asked to provide a 'very simple common name or label for each of the slides ... the most simple, obvious, direct sort of name that ordinary people would give for each scene.' In the second, subjects were required to complete sentences describing activities with appropriate names for settings, as in "The Kingstons furnished their _____ with furniture they built themselves." There was a high degree of consensus in both sets of responses, with the basic-level categories beach, mountain, city, and park (categories dictated to a large degree by choice of stimuli) being preferred even though more specific or more general terms would have been appropriate.

12.3   Lloyd et al.'s Research on Basic-Level Geographic Categories

The third study reviewed here was reported in Lloyd et al.'s (1996) article entitled "Basic-level geographic categories." In previous work by Rosch and others, such as work on folk taxonomies of plants and animals, folk taxonomies often appear to approximate scientific taxonomies, at least superficially. In contrast, Lloyd et al. propose that the common categories of administrative units in the United States (country-region-state-city-neighborhood) may be at the same basic level in a cognitive hierarchy of familiar categories and terms, with *place* as the superordinate category. In a departure from Rosch and other pervious workers, their model populates the subordinate category layer not with subclasses but with instances that are *particular cases* – such as the South, or Georgia, or Charleston. Lloyd et al. do not discuss the shift Rosch's theory in much detail. The shift between Roschian experimental methods and Lloyd et al.'s work is understandable, since geographic categories seem to seldom have many subclasses (e.g., kinds of mountains). However, this difference makes Lloyd et al.'s experimental results more difficult to compare with work for non-geographic categories.

Lloyd et al. tested 11 geographic terms, selected because they were reasonable answers to the question: "Where is your home?", and each subject was tested for just one term. *Place* was tested as the superordinate level, and country, region, state, city, and neighborhood were hypothesized to be at the basic level. The five subordinate level tests asked the subjects to list characteristics, activities, and parts for their *home* country, their home region, their home state, their home city, or their home neighborhood, listing that specific home place on each page of their response. The experimental protocol asked subjects to list as many "characteristics, activities or parts they associated with a particular geographic term," with 90 seconds for characteristics, 90 seconds for activities, and 90 seconds for parts (Lloyd et al. 1996, p. 187). Lloyd et al.'s main findings were that fewer characteristics, activities, and parts were listed for the superordinate term *place* than for the other groups. When averaged across subjects, results for the categories of country, region, state, city, and neighborhood were all extremely similar, suggesting indeed that they are all conceptual categories at a common cognitive level, presumably the basic level. They reported only minor differences in average numbers of characteristics, activities, or parts listed for, say, U.S. states in general, and for their home state. This appears to confirm our suspicion that testing specific instances (category members) is not the same as testing subclasses (subordinate categories).

13.   Experiments To Elicit Category Norms for Some Geographic Kinds

We propose to test a total of 17 categories using Battig and Montague's methods (Battig and Montague 1968). Six categories are new to this study (**bold faced type** in Table 1, below), and eleven are categories already tested by Battig and Montague, seven of which were somewhat geographic, and four non-geographic. For the eleven categories repeated from Battig and Montague, we will have a baseline for evaluating our results. Some non-geographic categories are included to make the objective of the study somewhat less obvious to the subjects.

Table 1: Categories to Be Tested in an Elicitation Task

---

1.   a precious stone
2.   a unit of distance
3.   a type of human dwelling
4.   a color
5.   **a kind of geographical feature**
6.   a country
7.   a crime
8.   a weather phenomenon
9.   a city

10. **a kind of water feature that would be shown on a map**
11. a bird
12. a natural earth formation
13. **a kind of geographic feature made by humans (not "natural")**
14. a US state
15. **a kind of human settlement (populated place)**
16. **a political entity**
17. **a kind of geographic object that typically has an indeterminate (fuzzy, graded, or uncertain) boundary**

For the categories repeated from Battig and Montague's study, we can only predict that our results will replicate theirs. In this section, we provide a brief discussion of anticipated results for the testing of the five new geographic categories we have added. If experimental results depart from these expectations, our representations of the naive theory of geographic kinds for the corresponding populations will have to be revised accordingly.

A kind of geographical feature. This will be the first truly-geographic category presented to the subjects. The main question here is whether natural or artificial features will appear on a greater number of response lists highly, also whether solid ground or water-related features will be more frequent. We predict mostly natural features, and in roughly the same order as for "a natural earth formation," but probably with features made by humans listed more frequently than for "a natural earth formation."

A kind of water feature that would be shown on a map. Based on the ontological work presented above, we predict that *lake* (or *sea*, depending on population) and *river* will rank first and second among kinds of water features. If experimental data deny this, we will have to re-think the ontological framework. We predict that other whole water bodies will be next (*pond*, *ocean*), then some kinds of watercourses (*stream*, *creek*), then parts of water bodies (*strait*, *bay*), and lastly some non-geographic water objects (*puddle*, *drop*), or borderline cases (*swimming pool*, *fish tank*). We will pay particular attention to the frequency ranks of standing water bodies, flowing watercourses, and parts of water bodies or watercourses, as well as ranking of specific entity types within these groups.

A kind of geographic feature made by humans (not "natural"). Categories such as *road* or *city* might be most highly ranked, but our draft ontology for geographic features made by people thus far only includes those features that pertain to artificial boundary demarcations. For this and the next category, results of the elicitation of norms will provide input to extensions of the ontology of these

kinds of things.

A kind of human settlement (populated place). *City* might be listed first here, but otherwise we have no prediction of the relative frequency order of *village*, *hamlet*, *town*, *borough*, *campground*, etc., and also have no prediction regarding the ranking of categories of extended settlement zone (such as *township*, *county*, or *state*) relative to conceptually point-like settlements.

A political entity. We tentatively anticipate an inverse ordering to that invoked by "a kind of human settlement." Results also will be compared to Lloyd et al.'s (1996) findings.

A kind of geographic object that typically has an indeterminate (fuzzy, graded, or uncertain) boundary. The ontology of geographic objects with indeterminate boundaries is in its infancy, although some of the material presented in Burrough and Frank (1995) will provide a valuable basis for the ontology. Responses to this category will provide further starting points for analysis.

14. Experiments Based on Rosch's Methods

15.1 Examples of categories

Rosch (1973) used Battig and Montague's norms as input to her "Experiment 3" on judgments about the internal structure of categories (Rosch 1973, pp. 130-134). She chose eight of Battig and Montague's categories for further research: *fruit, science, sport, bird, vehicle, crime, disease, vegetable*. From each of these selected categories, Rosch chose 6 instances across a range from very good to very peripheral members of the categories; she operationalized this criterion by choosing to test the instances with frequencies closest to 400, 150, 100, 50, 15, and under 5 in Battig and Montague (out of 442 subjects). She then gave subjects the categories, and for each, the list of 6 instances selected according to the above criteria, and asked subjects to rate, on a scale of 1 to 7, how good each example was of its category. She found that with few exceptions, instances given most frequently as exemplars (category norms in Battig and Montague's experiment) were also judged to be much better examples of the categories.

We will include a test based on Rosch's method, applied to Battig and Montague's category "A Natural Earth Formation."

Q: On a scale of 1 (excellent example) to 7 (poor example), please rate each of the following kinds of geographic objects regarding how good an example it is of "a natural earth formation":

		Excellent Example	2	3	4	5	6	Poor Example
a.	crater	1	2	3	4	5	6	7
b.	gully	1	2	3	4	5	6	7
c.	iceberg	1	2	3	4	5	6	7
d.	lake	1	2	3	4	5	6	7
e.	mountain	1	2	3	4	5	6	7
f.	river	1	2	3	4	5	6	7

This same protocol will be used also to test the category norms that result from the geographic categories included in our own experiment to elicit category norms.

15.2 Numbers of Attributes and Parts

Following Lloyd et al.'s (1996) use of one of Rosch's experimental protocols, we will ask subjects to list as many characteristics of some geographic categories as they can in 90 seconds, and other subjects to list as many parts of some geographic categories as they can in the same time period. Following Lloyd et al., each subject will be asked about only one category, and only about characteristics or parts (not both), at the beginning of a test that includes other questions. For comparison with Lloyd et al., we will include city in the set of terms tested. The others tested will be *lake*, *pond*, *bay*, *river*, *hill*, *mountain*, and perhaps others. We hypothesize that these natural geographic categories will be thought to have many characteristics and few parts.

15.   Questions on the Nature of Boundaries

We have hypotheses to the effect that geographic entities are associated with a distinct cognitive ontology in part due to special features of their boundaries. We will test whether fiat and bona fide boundaries are commonly considered to be different, and in what ways. We also will test whether crisp and indistinct boundaries are cognitively distinct. Sample questions that address this point include:

Q. If an island is divided into two political entities, how is the boundary between the political entities similar to the boundary of the island itself, and how is it different?

Q. In what ways are the boundary of a country and the boundary of an apple similar, and in what ways are they different?

Q. List some ways in which the edge of a wetland differs from the edge of a park.

Q. Who do you think owns the boundary between two adjacent land parcels in the area of your

home town?

  a. the owner of the oldest parcel
  b. the person who has owned one of the parcels the longest
  c. the two parcel owners each own their half of the boundary
  d. the boundary is jointly owned by both parcel owners
  e. the boundary has no owner

The responses to each of these questions will determine further questions to be asked in later phases.

16.   **Tests Related to Definitions**

In these tests, subjects would be asked to select the best definition of some water-related terms, in a multiple choice format:

  Q. Which of the following do you think is the *best* definition of a "lake":

  a. a large inland body of water
  b. a closed loop formed by a shoreline, with a water surface inside it
  c. water contained within a predominantly natural shoreline that exhibits no appreciable current
  d. an extent of water larger and deeper than a pond
  e. a part of the earth's surface, other than the ocean, covered by still water
  f. a depression in the earth's surface that is normally filled with water
  g. a large inland natural sheet of water
  h. a large inland body of fresh water or salt water.

Similar questions will ask subjects to choose definitions for *river*, *pond*, *bay*, and other water features.

17.   Conclusion: On Empirical Ontology

Empirical ontology as we conceive it will involve two complementary research methods: ontological work in the traditional sense (largely deductive, introspective, and formal) and research with human subjects (empirical, inductive). Ontological theories will be used as starting point for the formulation of experimental protocols designed to establish their degree of fit with corresponding systems of beliefs embodied in human cognitive systems. Analysis of data from human subjects will then be examined and generalized to produce compensating adjustments in ontological theories, which will in general lead to further rounds of empirical testing. (Compare the interplay of formal modeling and empirical research in Mark and Egenhofer 1994, 1994a, 1995; Egenhofer and Mark 1995.)

The ontological framework to be tested is a multi-leveled construction, involving not only mereology and topology, but also theories of spatial location and qualitative geometry, theories of fiat and bona fide boundaries and theories of vagueness and indeterminacy, along the lines of Smith

and Varzi (1997). In addition to the philosophical literature, another source of ontological theories is work in the artificial intelligence field (see for example Hayes 1985). In constructing software tools for merging large databases, it has proved fruitful to develop common ontologies in terms of which divergent bodies of data derived from different sources can be unified together into more compact systems. Ontological engineering of this sort was pioneered by Tom Gruber and his colleagues in Stanford, and a summary of recent work is represented in Guarino (1998). For spatial applications see Stock (ed.) 1997, and especially the contribution by Frank.

Much of the A.I. work on ontology recalls earlier investigations by analytic philosophers on the ontological commitments of scientific theories. (See Quine 1953) It goes beyond these, however, in that it seeks to reconstruct in their entirety the ontological theories embodied in given information systems and to put refined and simplified versions of these reconstructed theories to practical purposes within the information systems domain. The work described here is focused on the ontological theories embodied in human cognition, and it seeks to reconstruct these with the help of empirical testing. Note that this work is distinguished from investigations in epistemology: we are only incidentally concerned with the evaluation of geographic knowledge and with the questions pertaining to the ways in which human subjects come to know geographic categories.

The formal-ontological theory projected here will be developed axiomatically using the resources of first-order predicate logic. This will ensure ready translatability into the languages such as *ontolingua* (Gruber 1993) and into other ontology-based frameworks for database translation and knowledge interchange standardization. The axiomatization will embody a syntactic distinction between substantial and accidental predications, the former coding categories in a way which can allow representations of distinctions such as that between base- and non-base-level categories, dependent and independent categories, and so on. Finally, the axiomatization must enable us to distinguish, at least in principle, between geographic categories employed within a given culture and universal geographic categories which all systems of geographic categories share in common. That is to say, our abstract ontological framework must include at least the possibility of coding both culture-specific and universal features of human ontological belief-systems. One not inconsiderable benefit of the methodology here described turns on the fact that, as we hypothesize, strictly formal investigations in ontology may have important things to tell us about the universal constraints which all systems of geographic categories must satisfy.

References

American Heritage Electronic Dictionary 1992 Third Edition, Boston: Houghton Mifflin.

Asher, N., and Vieu, L. 1995 "Toward a Geometry of Common Sense: A Semantics and a Complete Axiomatization of Mereotopology," *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, San Mateo, CA: Morgan Kaufmann, pp. 846-52.

Battig, W. F., and Montague, W. E., 1968 "Category Norms for Verbal Items in 56 categories: A Replication and Extension of the Connecticut Norms," *Journal of Experimental Psychology* Monograph, 80, No. 3, Part 2, pp. 1-46.

Bennett, J. 1988 *Events and Their Names*, Indianapolis: Hackett.

Brogaard, Berit O., Peuquet, Donna and Smith, Barry 1999 *Objects and Fields. Report on the Specialist Meeting of Varenius Research Initiative*, Buffalo/Santa Barbara/Maine: National Center for Geographic Information and Analysis.

Burrough, P., and Frank, A. U., (editors) 1995. *Geographic Objects with Indeterminate Boundaries*. London, Taylor and Francis.

Casati, R., and Varzi, A. C. 1994 *Holes and Other Superficialities*, Cambridge, MA, and London: MIT Press (Bradford Books).

Casati, R., and Varzi, A. C. 1996 "The Structure of Spatial Location," *Philosophical Studies* 82, 205-239.

Casati, R., Smith, B., and Varzi, A., 1998 "Ontological Tools for Geographic Representation," in Guarino (ed.), 77-85.

Cohn, A. G., and Gotts, N. M., 1994 "Spatial Regions with Undetermined Boundaries," *Proceedings of the Second ACM Workshop on Advances in Geographic Information Systems*, 52-59.

Cohn, A. G., Bennett, B., Gooday, J., and Gotts, N. M., 1997 "Qualitative Spatial Representation and Reasoning with the Region Connection Calculus," *Geoinformatica*, 1(3), 1-42.

Cohn, A. G., Bennett, B., Gooday, J., and Gotts, N. M., 1997a "Representing and reasoning with qualitative spatial relations about regions," in Stock (ed.), 97-134.

Egenhofer, M. J., and Mark, D. M., 1995 "Modeling Conceptual Neighborhoods of Topological Relations," *International Journal of GIS*, 9, No. 5, pp. 555-565.

Egenhofer, M. J., and Mark, D. M., 1995a "Naive Geography," in Frank and Kuhn (eds.), 1-15.

Egenhofer, M., and Herring, J., 1991 "Categorizing Binary Topological Relationships Between

Regions, Lines, and Points in Geographic Databases," Department of Surveying Engineering, University of Maine, Orono, ME.

Estes, W. K. 1994 *Classification and Cognition*, New York/Oxford: Oxford University Press.

Frank, Andrew 1997 "Spatial Ontology," in Stock (ed.), 135–153.

Frank, A. U. and Kuhn, W. (eds.) 1995 *Spatial Information Theory: A Theoretical Basis for GIS,* Berlin: Springer-Verlag, Lecture Notes in Computer Sciences No. 988,

Gruber, T. R. 1993 "A Translation Approach to Portable Ontology Specifications," *Knowledge Acquisition*, 5(2),199-220.

Guarino, Nicola (ed.) 1998 *Formal Ontology in Information Systems*, Amsterdam, Oxford, Tokyo, Washington, DC: IOS Press (Frontiers in Artificial Intelligence and Applications).

Gunalik, I. B. and Friend, J. H. (eds.) 1966 *Webster's New World Dictionary of the American Language*, College Edition, Cleveland and New York: The World Publishing Company.

Hayes, P. 1985 "The Second Naive Physics Manifest," in: Hobbs and Moore (eds.), 1–36.

Hayes, P., 1985a "Naive Physics I: Ontology of Liquids," in: Hobbs and Moore (eds.), 71-108.

Hobbs, J. and Moore, R. (eds.), *Formal Theories of the Commonsense World*, Norwood, NJ: Ablex.

Keil, F. 1979 *Semantic and Conceptual Development: An Ontological Perspective,* Cambridge, MA: Harvard University Press.

Lloyd, R., Patton, D., and Cammack, R. 1996 "Basic-Level Geographic Categories," *Professional Geographer,* 48, 181–194.

Mark, D. M. 1993 "A Theoretical Framework for Extending the Set of Geographic Entity Types in the U.S. Spatial Data Transfer Standard (SDTS)," *Proceedings, GIS/LIS'93*, Minneapolis, November 1993, 2, pp. 475-483.

Mark, D. M., 1993a. "Toward a Theoretical Framework for Geographic Entity Types," in Frank, A. U., and Campari, I, editors, *Spatial Information Theory: A Theoretical Basis for GIS*, Berlin: Springer-Verlag, Lecture Notes in Computer Sciences No. 716, p. 270-283.

Mark, D. M., and Egenhofer, M. J., 1994 "Calibrating the Meanings of Spatial Predicates From Natural Language: Line-region Relations," *Proceedings, Spatial Data Handling 1994*, Vol. 1, 538-553.

Mark, D. M., and Egenhofer, M. J., 1994a "Modeling Spatial Relations Between Lines and Regions: Combining Formal Mathematical Models and Human Subjects Testing," *Cartography and Geographic Information Systems*, October 1994, 21, No. 4, 195-212.

Mark, D. M., and Egenhofer, M. J., 1995 "Topology of Prototypical Spatial Relations Between Lines and Regions in English and Spanish," *Proceedings, Auto Carto 12*, Charlotte, North Carolina, March 1995, pp. 245-254.

Mark, D. M., Egenhofer, M. J., and Hornsby, K., 1997 *Formal Models of Commonsense Geographic Worlds: Report on the Specialist Meeting of Research Initiative 21*, Santa Barbara, CA: National Center for Geographic Information and Analysis, Report 97-2.

Mark, D. M., and Frank, A. U., 1992. *NCGIA Initiative 2: Languages of Spatial Relations*. National Center for Geographic Information and Analysis, Santa Barbara, CA, Technical Report.

Mark, D. M., and Frank, A. U., 1996 "Experiential and Formal Models of Geographic Space,

 Environment and Planning, B, 23, pp. 3-24.

Mark, D. M., Comas, D., Egenhofer, M. J., Freundschuh, S. M., Gould, M. D., and Nunes, J., 1995. "Evaluating and Refining Computational Models of Spatial Relations Through Cross-Linguistic Human-Subjects Testing," in Frank and Kuhn (eds.), 553-568.

Mostowski, M., 1983 "Similarities and Topology," *Studies in Logic, Grammar and Rhetoric*, 3, 106-119.

Petitot, Jean 1995 "Morphodynamics and Attractor Syntax: Constituency in Visual Perception and Cognitive Grammar," in Robert F. Port and Tim van Gelder, eds., *Mind as Motion. Explorations in the Dynamics of Cognition*, Cambridge, MA and London: MIT Press, 228–281.

Quine, W. V. O. 1953 "On What There Is," *From a Logical Point of View*, Cambridge, MA: Harvard University Press.

Rosch, E. 1973 "On the Internal Structure of Perceptual and Semantic Categories," in T. E. Moore (ed.), *Cognitive Development and the Acquisition of Language*, New York, Academic Press.

Rosch, E., 1978 "Principles of Categorization," in E. Rosch and B. B. Lloyd (eds.) *Cognition and Categorization*, Hillsdale, NJ: Erlbaum.

Simons, P. M. 1987 *Parts. An Essay in Ontology*, Oxford: Clarendon Press.

Smith, B., 1994 "Fiat Objects," in N. Guarino, L. Vieu and S. Pribbenow (eds.), *Parts and Wholes: Conceptual Part-Whole Relations and Formal Mereology, 11th European Conference on Artificial Intelligence, Amsterdam, 8 August 1994*, Amsterdam: European Coordinating Committee for Artificial Intelligence, 15-23.

Smith, B., 1995 "On Drawing Lines on a Map," in Frank and Kuhn (eds.), 475-484.

Smith, B., 1996 "Mereotopology: A Theory of Parts and Boundaries," *Data and Knowledge Engineering*, 20, 287-303.

Smith, B. 1997 "The Cognitive Geometry of War", in Peter Koller and Klaus Puhl (eds.), *Current Issues in Political Philosophy: Justice in Society and World Order,* Vienna: Hölder-Pichler-Tempsky, 394–403.

Smith, B., and Varzi, A., 1997 "Fiat and Bona Fide Boundaries: An Essay on the Foundations of Geography," in S. C. Hirtle and A. U. Frank (eds.), *Spatial Information Theory. International Conference COSIT '97. Laurel Highlands, Pennsylvania, October 1997* (Lecture Notes in Computer Science 1329), Berlin/New York: Springer Verlag, 103-119. Revised version forthcoming as "Fiat and Bona Fide Boundaries", *Philosophy and Phenomenological Research.*

Smith, B., and Varzi, A., (in press) "The Niche," *Noûs*, URL: http://wings.buffalo.edu/academic/ department/philosophy/faculty/smith/articles/niches.html.

Stock, Oliviero (ed.), 1997 *Spatial and Temporal Reasoning*, Oliviero Stock (Ed.), Kluwer Publishing Company,

Tversky, B. and Hemenway, K. 1983 "Categories of Environmental Scenes," *Cognitive Psychology,* 15, 121-149.