ELSEVIER

Fuzzy Sets and Systems 113 (2000) 7–18

FUZZY sets and systems

www.elsevier.com/locate/fss

# Sorites paradox and vague geographies

Peter Fisher

*Department of Geography, University of Leicester, Leicester LE1 7RH, UK*

Received November 1998

**Abstract**

The sorites paradox is ranked among the top five paradoxes of philosophy (Sainsbury, Paradoxes, 2nd ed., Cambridge University Press, Cambridge, 1995). It is simply stated as 'what is a heap'. Deriving from the paradox is a definition of vagueness, which is contrary to the Boolean concept of the world implicit in much geographical teaching and thought, and the representation of geographical information in modern geographical information system. The argument of Sorites Paradox is suggested as a test of whether a concept is vague. If that concept is sorites susceptible, then it should be modelled as a vague concept, otherwise a Boolean model may be appropriate. The recognition of whether or not a particular concept is sorites susceptible does not have to influence the methods of analysis. It should merely inform the interpretation, and the investigator and reader should be aware that the outcome of the analysis is only one of a set of possible outcomes, which depends on how the vague concept is crispened. Furthermore, it is argued here that very many geographical phenomena (relations, objects and processes) can be shown to be sorites susceptible, and so vague, both generically and genetically. Vagueness can be addressed by multi-valued logic and applications of fuzzy set theory (the most common method of implementing multi-valued logic) to geography are reviewed. A formal recognition of vagueness in geographical phenomena is long overdue, and should be welcome in geographical analysis and, certainly, in geographical information systems. 

*Keywords:* Fuzzy sets; Sorites paradox; Vagueness; Vague objects; Vague relations; Geographical information systems

## 1. Introduction

*Do you know the difference between 'village' and 'hamlet'? Surprisingly few people do, but it's quite simple really: one is a place where people live and the other is a play by Shakespeare.*

Bill Bryson, 1995, *Notes from a Small Island*

The Sorites Paradox (the paradox of the heap) appears to have been originally formulated as a philosophical paradox by Eubulides of Miletus [3,10,87], although arguments of the same kind can be dated earlier and even appear in the Bible (Genesis 18: 23–33). The paradox is judged to be among the most profound and important of all those known to philosophers [77]. It is outlined below, and can be paralleled in very many and varied geographical concepts and objects. Because the paradox is at the core of the definition of vagueness in philosophy and logic, it is also fundamental to an appeal for a formal approach to vagueness in geography.

Vagueness is endemic in the human condition [87]. It is in our view and understanding of everything

* Tel.: 0116-252-3839; fax: 0116-252-3854.
*E-mail address:* pff1@le.ac.uk (P. Fisher)

0165-0114/00/$ - see front matter 
PII: S0165-0114(99)00009-3

around us, and, most profoundly, embedded in our natural language. Logical positivists have claimed that natural languages which use poorly defined terms and concepts are equivalent to nonsense [2], but that is to preclude most of the vocabulary of everyday human language from having meaning. Such an interpretation dismisses the richness of natural language, and ignores the importance of understanding the vagueness, which characterises it. The concept of vagueness can be extended from the vocabulary of every-day language to that of technical language and to objects in the real world [76,87].

This article starts by examining the sorites paradox in its original formulation, and uses it to define vagueness. The paradox is then presented as a test of whether any concept is vague. The concept of vagueness is then explored in geographical language and objects, and the case is made that vagueness is endemic in geographical thinking, and in geographical information. The argument is illustrated by a wide collection of examples of vague geographical concepts from many geographical disciplines. It is argued that vagueness is an inherent property of geographical data, and that ignoring it is to strip away the essence of much of those data. The significance of this conclusion for modern geographical information systems and the interaction between geography and approaches to systematizing the description of vagueness through fuzzy set theory is reviewed.

## 2. The sorites paradox

The sorites paradox is easily stated:

Is one grain of sand a heap of sand?
The answer to this simple question is clearly NO.
If a second grain is added to the first, is there a heap?
Again the answer is NO.
If a third grain is added, is there a heap?
For a third time the answer is NO.

The argument so far is uncontroversial. From it, however, we can conclude a general form of argument:

If there are $n$ grains, but no heap, then adding one grain to make $n+1$ grains will not make a heap.

By repeated application of this premise we can see that as $n$ increases to a large number, the addition of a single grain still does not change a non-heap into a heap, and this is one crux of the argument. The addition of one single grain can *never* turn a collection of grains into a heap. Thus, we may have any number (millions) of grains of sand and not have a heap. Similarly, the addition of one hair cannot change a bald man to a hirsute one, and the increment of one unit of hue towards blue can never change the colour viewed from red to blue (two more of the classic lines of argument related to sorites paradox; [87]; see also [88]). This conclusion is paradoxical. We have an initial condition that is true; one grain is not a heap. We have a premise, which is apparently true: for any value of $n$, adding 1 grain will not turn a non-heap to a heap. At the end of a repeated application of the premise we have the false conclusion that a collection of a million grains (for example) is not a heap. Having a correct sequence of premises, which reach a false conclusion, is a paradox [77], and this one is named after the Greek word for heap (*soros*) ([77], pp. 23–51; [78,87]). It should be noted that this is a paradox, not a fallacy. A fallacy is similar in that it reaches a false conclusion, but it is based on either a false initial premise or erroneous logic [21]. The sorites paradox seems to have neither.

More formally, if $H'$ is the concept of not-a-heap, and there is a set $[a_1, a_2, \ldots, a_n]$ of number of grains, $1, 2, \ldots, n$, then

$$\begin{array}{l} H'_{a_1} \\ H'_{a_1} \to H'_{a_2} \\ H'_{a_2} \to H'_{a_3} \\ \vdots \\ H'_{a_{n-1}} \to H'_{a_n} \\ \hline H'_{a_n} \end{array}$$

where $p \to q$ means if $p$ then $q$, and $H'_{a_1}$ means that one grain is not-a-heap.

The argument also works in reverse:

If we have a heap of sand, and remove one grain.
Is it a non-heap?
If we remove another grain, is it now a non-heap?
...

We do not lose the heap so long as one grain remains. Again, more formally, for the concept $H$ (there is a heap), and assuming that $n$ is sufficiently large that there is no doubt but that initially there is a heap:

$$\begin{array}{l} H_{a_n} \\ H_{a_n} \rightarrow H_{a_{n-1}} \\ H_{a_{n-1}} \rightarrow H_{a_{n-2}} \\ \vdots \\ H_{a_2} \rightarrow H_{a_1} \\ \hline H_{a_1} \end{array}$$

As before, we have a true initial premise, a logical inference (if $p$ then $q$) and its repeated application. Again there is no precise point when the removal of a single grain turns the heap into a non-heap. This reverse inference chain need not end at 1 grain. If 1 grain is a heap, and one is removed, given the general form of the argument, then a heap will remain! How can no grains of sand be a heap? Equally we know that 1 grain of sand is not a heap, and that millions of grains do form a heap. The form of the argument is correct. The conclusion, however, is false.

The main point of the argument is that a heap is a poorly defined concept. While it is the sort of word we use in describing things the whole time, it is not precise, and never can be. It is inherently vague in its meaning and application. A logical positivist (or pragmatic) view of the argument might suggest that a particular value should be defined as a threshold of heapness. When $n =$ threshold, then the question “is this a heap” becomes true. Indeed, this approach has been adopted by many in discussion of the paradox, but such a definitional threshold of a heap strips the word of its inherent meaning, which is steeped in its vagueness. The vagueness is not an inconvenient imprecision of expression, but a fundamental property of anything we might term a heap. If we wished to be precise, we would take the time to count or weigh the grains, to give a precise quantification of the amount of sand.

Paradoxes attract suggestions of solutions (or proofs that they are fallacies). There are a number of possible such suggestions for sorites paradox.

1. *The initial premise is false*. In many instances this is not the case; one grain by itself is not a heap.
2. *The process of reasoning is false*. The reasoning process is called *modus ponens* which essentially states that if $p$ then $q$, and so given $p$, $q$. It is one of the most fundamental clauses of logical inference, and in sorites paradox it is applied repeatedly. It is possible that its repeated application makes it invalid, but that would be against any rules of logic; if it is true, then it is always true.
3. *The Law of Excluded Middle*. It is implicit in the argument proposed [36] that for any $x$, either $x$ is $F$ or $x$ is not $F$. In other words, nothing can be a heap, and, at the same time, a non-heap. If this were allowed, then we could admit that at 25 grains (for example) the collection of grains was to some extent a heap, not a complete, true heap, but also not a true non-heap. If the Law of Excluded Middle is discarded, then it is possible to address the sorites paradox, because for some number of sized grains there are actually two questions to which the answer is maybe, or even yes: is it a heap *and* is it a non-heap. A semantic approach to the paradox regards *degrees of truth* as an acceptable solution to the sorites paradox [77], but an epistemic approach to vagueness does not [87]. On the other hand, even if degrees of truth are only a partial solution, they can be implemented within a scientific context, and can form the basis of a vague geography.

Sorites paradox is usually used in philosophy to address vagueness of language, and a long running argument has arisen over whether or not there are vague objects. The argument revolves around whether or not an object is vague or only its conceptualization. It might be thought that a heap is an object, and so the paradox itself is defined and named in terms of the existence or otherwise of a vague object. Perhaps only the concept of the heap is vague, while the object itself is sharp since we could possibly measure the number of grains. In spite of arguments such as that of Evans [22] who argues that vague objects are inconsistent, the majority opinion amongst recent writers is that there are indeed such objects [9,76,77,87].

At least two principal views of the root cause of vagueness can be identified. The first is semantic and seems to view vagueness as principally a matter of degree. There is a degree to which a number of sand grains is or is not a heap. Sometimes the degree is certain, sometimes it is vague [77]. The second is epistemic vagueness where essentially the boundary between a heap and a non-heap is very real. For any individual at any time, there is a clear conception of what constitutes a heap and what does not: a crisp boundary can be identified between the two. The problem of vagueness arises because the individual does not retain a single opinion, rather the opinion fluctuates with time, with further information, outside influences, etc. The individual can therefore identify very precisely a heap as opposed to a non-heap, but the boundary is a moving target [87]. There may be little practical difference between the two versions of vagueness, but they are profoundly different understandings of the phenomenon.

The problem of vagueness (of whichever flavour) is that it undermines Aristotelian (traditional) logic, which is based on Boolean conditions and such principles as the Law of Excluded Middle and *modus ponens*. Much of the writing on approaches to vagueness are attempts either to match vagueness with such logic, which seems doomed to failure, or to redefine logical concepts to accommodate vagueness. Neither has been wholly successful [87].

## 3. Sorites tests for vague geographical concepts

Sorites paradox can be used as a test of whether a concept is vague, and it is used in the following sections to show that concepts in many areas of geography are vague.

### 3.1. Geographical relations

Perhaps the most common spatial concept of a sorites paradox is proximity. We commonly refer to one location as being near or close to another (as well as by other possible synonyms and antonyms). What do we mean by near? If we assume that proximity can be measured by distance away (which is not necessarily the case), then if we are at the location of object A (if it is a point, or at the edge of its extent if it is a polygon, or line), then if we move one metre away are we near it? If we move a second metre away, are we near? In the light of earlier discussion, we are always near! Near and similar geographical concepts are sorites susceptible, and so vague.

Directional concepts of relative location are also vague, but they also suffer from ambiguity. Ambiguity is confusion among concepts which have the same name, but more than one precise definition. North in its technical sense means either that one location lies directly north of another, or that one object is closer to the North Pole than another. In the first definition the two objects are on the same line of longitude, while in the second they could be 179.99° latitude apart, but 0.01° longitude closer to the north pole. Further confusion may come from the frame of reference; whether north is with respect to magnetic, true or grid north. The confusion here is known as *ambiguity*, and is easily clarified by stating the frame of reference, and the alternative definition in use in any situation.

The second definition of north as given above can be given a more general meaning: that one object lies somewhere vaguely to the north of another as opposed to south, east or west of it (perhaps between northeast and northwest, but not necessarily). This meaning is the colloquial use, employed in many every-day situations of giving directions, and is indeed sorites susceptible. If one object is precisely north of another, then it is to the north of it. If it is one degree to the east (or west) of north is it no longer to the north? This vague concept reflects the true human use of *north of* more precisely than does any precise definition.

### 3.2. Geographical objects

The urban–rural divide is fundamental to much human geography. Let $H$ be the concept of an urban area, $H'$ the concept of a rural area, and $a$ the number of people or residential buildings. Both the forward and reverse sorites arguments given above work, so either cities do not exist or rural areas do not, and possibly both. Actually defining a clear measure of the urban–rural divide is much harder. It can perhaps best be argued on either population density, house density, or absolute count in contiguous high density areas, but the measure decided upon makes no difference to the sorites paradox, because all measures are susceptible. Naturally a single measure is itself inappropriate, but any multivariate combination of measures does nothing to improve the situation. If a threshold is set for population density *and* number of service outlets *and* contiguity of high density areas then if a candidate for being a city is one item over the threshold in any measure, then does removing one of any make it necessarily a rural location? Introducing an intergrade zone perhaps known as suburban land use may ease the problem, but definition of suburban-urban and suburban-rural boundaries then become a matter of concern.

In colloquial use, the city is synonymous with a densely settled urban area. A more restricted definition of a city is in the legal and political context. Here an urban area is a city because it is defined as such if the seat of one of a number of particular services, and frequently several occur in the city. The services are either legal (county courts, for example), ecclesiastic (the seat of the diocese) or vernacular (county government). This distinction may seem to hark back to the distinction between formal and functional regions, but this may be more apparent than real. Some regions, both formal and functional, are sorites susceptible, while others of both types are not.

If an urban area cannot itself be easily distinguished, then its spatial extent is also problematic. A principal property of a city is its population. If the extent of the city is vague, then surely its population is also a matter for speculation. Yes, the number of people who live within the legal definition of a city, the city limit, can be determined. How many people are perceived as living in the city depends on the perceived extent of the city, and does not have to relate to the number of people who actually live within that area.

A simple construct from biogeography is woodland. How many trees are needed for an area to be classed as woodland? What is the amount of canopy closure required? Does one less tree (or one less per cent of canopy) make it non-woodland, a scrubland perhaps? In reverse, does one more make it woodland? Indeed, Moraczewski [61] has reviewed a host of plant community descriptions showing that they are all vague in their definition (see also [69]). Furthermore, when is woodland (assuming we can define it) oak woodland; does the change from 40% oak to 41% oak in the tree species per square kilometre make it so?

Many climatic and meteorological phenomena are also sorites susceptible. Does one drop of rain constitute a rainstorm? From a sorites type argument, we can derive the reassuring proposition (if you live in Britain) that rainstorms never occur! Equally, either all climates are hot or all are cold, based simply on the vagaries of how the paradox is posed.

We can also ask when is a hill a mountain? This version of the paradox has been given hilarious treatment in the recent film and book *The Englishman who went up a hill and came down a mountain* [60]. The film's comedy is derived in large part from the artificial sharpening of the boundary condition between a hill and a mountain, based solely on an elevation of 1000 ft, which motivates a whole village to a frenzy of activity for a week [30]. Not only is elevation on its own insufficient to define a mountain [81], but even if it were sufficient the boundary condition for a mountain as compared with a non-mountain cannot be determined as a generality. Indeed, as Sainsbury [76,77, p. 25] notes the geographical extent of a mountain is hard to define. It is not only scale-dependent [89], but vague in any absolute sense; indeed, any precise definition would be trivial. The Munroes of Scotland (broadly the Scottish peaks over 3000 feet) provide such an example. They are an achievement target for walkers and climbers [79].

In a different context, the Metropolitan Statistical Areas (MSA) of the US Census are those counties which contain a city with a population over 50 000 [4,91; 64, p. 68]. Furthermore, the area must have a metropolitan character as defined by having more than 75% non-agricultural labour, and other criteria in a similar vein. Are these thresholds meaningful in respect of the *process* being studied, urbanization and demography? Is a city with 49 999 people or 74% non-agricultural population fundamentally different from one classed as an SMSA? The threshold is used as a basis for urban development funding, and so it is, in reality, a sharpening of a vague concept (Metropolitan areas) for the sake of convenience in policy making. It is the setting of a threshold of a measurable property to classify a location into a class which has important and very real resource implications for people.

Many further geographical phenomena have this same problem of definition which relates to the sorites paradox. When is a sand dune a sand dune, and not a ripple? When is an ocean wave a wave? When is a soil an alfisol, and not an ultisol? The list is very long. How high does the temperature (and related climatic variables) have to be for it to be a hot climate? Generally, the object corresponds to the geographical individual [40, p. 216], and the basic approach to individualization (identification of the individual or object as a basis for analysis) described by Harvey seems to rely on setting thresholds. Any such threshold is necessarily sorites susceptible, whether it is defined in univariate or multivariate space.

### 3.3. Geographical processes

The description and definition of the geographic object is merely a means to an end, and not the end in itself. The geographer is concerned by the patterns of phenomena and the objects produced by those phenomena. Indeed, measurement and discourse in geography (as in other sciences) is commonly one means to assisting an understanding of process. It is not the soil type at a location which is important, so much as understanding why it is there; not so much the city itself as the processes which occur within the city and between the people who live there. In many instances, geographers have a relatively well developed concept of the broad processes of interest (human migration, soil formation, drumlin formation), but in almost all cases the details are less well understood. The details of the processes are usually more complex than a certain set of observations and arguments will allow us to examine. Just as in the definition of the spatial extent of geographical objects, and the class of objects, so too the concept of process is also subject to well-defined central concepts surrounded by broad zones of transition and vagueness.

In short, many geographical relations and objects can be seen to be vague, and implicit in their vagueness is the fact that their associated properties are vague. Such vagueness may be either endemic in the relation or in the concept, but has been ignored in the development of computer-based geographical databases (as well as most other databases). Interestingly, many of the luminaries of vagueness have used geographical objects in their discussions [75–77,87].

## 4. Degrees of truth

A number of methods have been suggested which attempt to address the problem of vagueness. The simplest method is three-valued logic, which has made a small impact on geographical thought and mapping. Of more significant impact has been the application of many-valued logics, and particularly the development of fuzzy sets.

### 4.1. Many-valued logics

A number of different approaches have been taken to defining degrees of truth, mostly involving many valued logics [87]. The simplest method is three-valued logic. The earliest discussions of sorites paradox by ancient Greek philosopher involved the Stoics who, under advice from Chrysippus [10], refused to respond to all questions which identified a boundary condition. Thus they would respond to inquiry in the following way:

| | |
|---|---|
| Is one grain a heap? | No |
| Are two grains a heap? | No |
| … | |
| Are $x$ grains a heap? | No |
| Are $x+1$ grains a heap? | No answer |
| … | |
| Are $y$ grains a heap? | No answer |
| Are $y+1$ grains a heap? | Yes |
| Are $y+2$ grains a heap? | Yes |
| … | |

where $y > x$.

In this approach $x$ and $y$ are conservative estimates of the limiting conditions of the heap and of the non-heap. They are respectively well to the small side for a possible heap and well to the large side for a definite heap. Implicitly the response identifies a range of values for the size of the heap when it is a possible heap where the group of grains is both a heap and a non-heap. Therefore, the Law of Excluded Middle is dented with respect to the concepts of heap and non-heap, although it can also be seen as being preserved because the new condition is a third possibility. The situation becomes more problematic, however, because in common sense there is no reasons why $x$ and $y$ should be the same for the sorites and reverse sorites cases. Neither the threshold of definite heapness, nor of definite non-heapness, has to be the same when increasing the numbers of grains as when decreasing the number.

In the Stoic argument, a boundary condition is implicit but the essence of Stoic philosophy can be preserved by declining to answer all questions in this range [10,87]. A full three-valued logic was developed by Halldén [39] in his *Logic of Nonsense*, meaning the logic which includes neither true nor false: a logic which sets aside the Law of Excluded Middle.

In examining the possible errors in the point-in-polygon query, Blakemore [8] shows a case of three valued logic within GIS where a point is either determined to be in one polygon, the other polygon or to be indeterminate. Some recent writing in geographical topology has examined the possibility of three-valued logic [16,17], but there seem to be few cases of formal applications of this approach to geographical information, although it is implicit in much reporting of research. A formal approach to three-valued logic which has been widely implemented for data retrieval in computer systems is the concept of Rough Sets [65], although its relevance to geographical information has not been researched.

Continuum valued logics have been much more popular than three-valued logics. The simple reason for this is that if an object is hard to define, so too are the threshold values of the category "neither true nor false", although it should be noted that this problem is not avoided by continuum valued logics, it is only easier to ignore. The primary philosophical work on this is Lukasiewicz and Tarski [54], followed by Black [7]. The most influential study in recent years is undoubtedly the introduction of Fuzzy Sets to information science by Zadeh [92], which was itself directly preceded by the work of Kaplan and Schott [41].

### 4.2. Fuzzy set theory

Fuzzy sets and fuzzy logic were first introduced by Zadeh [92] who is reported as deliberately not using the term vagueness to avoid any suggestion of a controversial, possibly unacceptable heredity [45]. Fuzzy sets are a direct (but partial) implementation of the concepts of vagueness discussed above, which give the appearance of directly addressing the problems associated with sorites paradox, and so have the potential to address or help with distinguishing more usefully the geographical concepts reviewed above.

In essence fuzzy sets are a generalization of classic, Boolean sets attempting to accommodate vagueness in the set boundary. In classic set theory an object is defined as belonging or not belonging to a class or set, and this may be coded as a binary membership function having values of 1 or 0 (recognizing the Law of Excluded Middle). In fuzzy set theory, the binary membership is extended to an infinite-valued membership from 0 to 1. This can be shown to enable a logic, which is the extension of traditional binary set logic, allowing negation, union, intersection, etc. The full set of fuzzy set operations is large, and is covered in a number of recent textbooks [43,46], as well as numerous introductory articles [1,12,35,50,66,71]. Review of these operations is not the purpose of the current article.

There has been a long-running discussion between the probability theorists, on the one hand, and the fuzzy set theorists, on the other [93,44]. A recent paper by probabilists Laviolette and Seaman [49] provokes a number of comments both for [53] and against [5,20] their criticisms of fuzzy set theory. The objections listed by Laviolette and Seaman [49] are essentially to do with applications of fuzzy logic, which yield a decision. Arguably, any decision is a Boolean event, and so it can be seen as either the crispening of a fuzzy set or suitable for treatment by probability theory. Laviolette and Seaman [49] ignored, however, the point that fuzzy theory is about poorly defined, vague sets. Probability is about identifying whether or not an object belongs to a Boolean set. The frequently quoted response to fuzzy set theory (re-deployed by Laviolette and Seaman [49]) that if you are trying to decide whether someone is in the set of bald men you are asking the wrong question, is steeped in the logical positivist view that any vague concept is meaningless [2]. Their argument seems to be based on entrenched views, and misconceptions of fuzzy sets, and specifically mis-construes the attempt in fuzzy set theory to formalise an understanding of the vague term *bald*. In short, probability deals with problems related to lack of data, while fuzziness deals with lack of definition.

Probability and fuzzy set membership do share much common ground. First and foremost, they are evaluated on a scale from 0 to 1. Probability, however, has a pedigree which goes back many years, while fuzzy set theory only goes back just over 20 years, although the real roots are just as old as those of probability theory. The real problem in understanding the distinction lies in the fact that probability is a fundamental part of the training of modern scientists, and certainly of many geographers, while the discussion of vagueness and fuzzy sets is not.

Williamson [87, pp. 120–130] identifies a number of issues related to fuzzy set theory. Most relevant is that although fuzzy logic claims to address the Law of Excluded Middle, and to overturn some aspects

Table 1
Some recent research on the vague and fuzzy conceptualization of geographical information. Not all articles fit the heading simply, but the closest association is indicated

| | General arguments for vagueness | Fuzzy semantic relation model | Fuzzy semantic import model |
|---|---|---|---|
| Climate | | McBratney and Moore [59] | Leung [51] |
| Vegetation | Moraczewski [61,62] | Dale [18], Roberts [68–70] | |
| Soil/land evaluation | | McBratney and De Gruijter [58] | Burrough [12,13]<br>Burrough et al. [14]<br>Davidson et al. [19]<br>Lagacherie et al. [47]<br>Wang and Hall [84]<br>Wang et al. [85] |
| Remote sensing | | Fisher and Pathirana [28,29]<br>Foody [31–33], Foody et al. [34]<br>Robinson and Strahler [73]<br>Robinson and Thongs [74]<br>Wang [82], Wilkinson et al. [86] | |
| Landscape | Fisher and Wood [30] | Usery [81], Wood [89] | |
| Natural language | | Altman [1], Fisher and Orf [27]<br>Robinson [72], Wang [83] | |

of classic logic, it is in fact based on classic logic [87,92].

## 5. Geographical application of fuzzy sets

A suite of early papers in the geographical literature advocated fuzzy sets especially as a basis for behavioural geography and for geographical decision making [35,50,66,67]. More recent treatments of similar topics show a continuing and developing interest in this same basic area [52,63]. In the present article, however, we are more interested in the conceptual models of geographical information and the role of fuzzy sets. Although not previously acknowledged, sorites paradox serves as a test of fuzziness in all the applications reviewed in this section because the concepts fuzzy sets are applied to are sorites susceptible.

Robinson [71] recognizes two methods for deriving fuzzy set memberships.

1. The Similarity Relation Model is based on cluster analysis and numerical taxonomy. It involves searching a dataset of measurements for pattern, and automated estimation of the membership. The fuzzy *c*-means classifier [6] is an early method for this, but more recently fuzzy-neural networks are being used [33,86].
2. The Semantic Import Model is where some form of expert or empirical model is devised to specify a formula for the membership function. This approach is that originally suggested by Zadeh [92] and adopted and reviewed widely since [12,24].

These two approaches are reflected in many geographical applications of fuzzy set theory (Table 1), and both can be justified by recourse to the sorites paradox. Interestingly, methods which are data-rich such as remote sensing seem to employ the Semantic Relation Model for determining memberships, while in other applications where data is more sparse such as in soil science and land evaluation, the Semantic Import Model has been employed. Most articles on fuzzy sets in geography do express their arguments in terms of vagueness, but few identify the origins of these ideas.

The apparent dominance of remote sensing in the literature cited in Table 1 is due to the specific reason that fuzzy set theory has been argued to address the so-called mixed pixel problem [26,73] which is a persistent problem limiting the use of remotely sensed data. Other applications are mainly motivated by problems in the specific scientific domain (soil science, land evaluation, climatology, and vegetation science). The mapping of phenomena (soils in particular) have, however, informed some applications and the importance of this has been increased by the automation of mapping and use of the geographical data within GIS.

In the area of natural language, research in fuzzy set theory has particularly been motivated by the wish to improve interfaces between GIS and the general public. This user group lives in a world steeped in vagueness where they function effectively, and they think about geography and space as vague concepts. It is essential that the geographical databases being created so rapidly as part of the expansion of GIS should use the same vagueness in the user interaction. In spite of considerable research into accuracy [37] and quality [38] in geographical databases the idea that the objects being stored are not well defined still seems to be ignored by both system developers and many academic researchers. Three- and greater-valued set theories provide a framework, in which vagueness can not only be developed and implemented, but also analysed as a basis for exploration and explanation.

## 6. Conclusion

In this paper vagueness has been traced back to its philosophical roots in the sorites paradox. The nature of the sorites argument has been shown, and it has been suggested as a useful test of whether an object or concept is vague. Many geographical phenomena such as proximity and directional relations and many geographical objects including cities, woodlands, mountains and rainstorms, among many others, have been shown to be sorites susceptible. That being the case, all these geographical objects and relations are themselves vague. At a semantic level, vagueness may be treated by recourse to many-valued logic, and fuzzy set theory is one method adopted by increasing numbers of information scientists and geographers in the use and application of geographical information. It is acknowledged here that fuzzy set theory is not a complete answer to the question of vagueness, but it is convenient because it can be implemented.

Geographical databases and geographical analyses should give a wider recognition to the vagueness of the concepts they are storing and analysing. Above all, the general public who may be interacting with the systems live in a world steeped in vagueness where they function effectively, and are used to thinking about geography and space as a vague concept. It is essential that the geographical databases being created so rapidly as part of the expansion of GIS should use the same vagueness in the user interaction. In spite of considerable research into accuracy [37] and quality [38] in geographical databases the concept that the objects being stored are not well defined still seems to be ignored by both system developers and many academic researchers. Fuzzy set theory provides a framework in which vagueness cannot only be developed and implemented, but can be analysed and sustained as a basis for exploration and explanation.

Vagueness is in the world we occupy, as human beings, and it is an essential part of how we perceive and understand that world. It seems necessary that analyses of geographical phenomena should test whether or not the phenomena are vague, using the sorites argument. If it is found that the phenomenon or the way it is being measured is vague, then that should be acknowledged. The vague concept can be sharpened or crispened, and the investigator can recognise that the outcome of analysis is only one of a set of possible outcomes dependent on the thresholds chosen. Alternatively, deliberate attempts may be made to address the vagueness in the concepts analysed, which may be based on three- or many-valued logics such as fuzzy set theory. Vagueness is a necessary part of the human experience of geography. The belief is stated here that analysis of geographical data should recognise this, not implicitly but directly.

## Acknowledgements

The opinions expressed in this paper are my own, but they owe much to discussion and development over many years, and particularly with John Dalrymple, Peter Burrough, Mike Goodchild, Vince Robinson, Dave Unwin, and Jo Wood. Peter Burrough, Mike Goodchild, Mark Sainsbury, and Dave Unwin commented on an earlier version of the manuscript, keeping some of the opinion under control and facts correct. The contribution of anonymous referees is also acknowledged. I thank all for their encouragement and opinions.

## References


[1] D. Altman, Fuzzy set theoretic approaches for handling imprecision in spatial analysis, Internat. J. Geograph. Inform. Systems 8 (1994) 271–289.

[2] A.J. Ayer, Language, Truth and Logic, Gollanz, London, 1936.

[3] J. Barnes, Medicine experience and logic, in: J. Barnes, J. Brunschwig, M. Burnyeat, M. Schofield (Eds.), Science and Speculation, Cambridge University Press, Cambridge, 1982, pp. 24–68.

[4] R. Barr, A comparison of aspects of the US and UK censuses of population, Trans. GIS 1 (1996) 49–60.

[5] J. Bezdek, Fuzziness vs. probability: again (!?), IEEE Trans. Fuzzy Systems 2 (1994) 1–3.

[6] J.C. Bezdek, R. Ehrlich, W. Full, FCM: the fuzzy *c*-means clustering algorithm, Comput. Geosci. 10 (1984) 191–203.

[7] M. Black, Vagueness: an exercise in logical analysis, Philos. Sci. 4 (1937) 427–455.

[8] M.J. Blakemore, Generalization and error in spatial databases, Cartographica (1984) 21131–21139.

[9] J.A. Burgess, The sorites paradox and higher-order vagueness, Synthese 85 (1990) 417–474.

[10] M.F. Burnyeat, Gods and heaps, in: M. Schofield, M.C. Nussbaum (Eds.), Language and Logos, Cambridge University Press, Cambridge, 1982, pp. 315–338.

[11] P.A. Burrough, Principles of Geographical Information Systems for Land Resources Assessment, Oxford University Press, Oxford, 1986.

[12] P.A. Burrough, Fuzzy mathematical methods for soil survey and land evaluation, J. Soil Sci. 40 (1989) 477–492.

[13] P.A. Burrough, Natural objects with indeterminate boundaries, in: P.A. Burrough, A. Frank (Eds.), Spatial Conceptual Models for Geographic Objects with Undetermined Boundaries, Taylor & Francis, London, 1996, pp. 3–28.

[14] P.A. Burrough, R.A. MacMillan, W. van Deursen, Fuzzy classification methods for determining land suitability from soil profile observations and topography, J. Soil Sci. 43 (1992) 193–210.

[15] J.B. Campbell, W.J. Edmonds, The missing geographic dimension of soil taxonomy, Ann. Assoc. Amer. Geograph. 74 (1984) 83–97.

[16] E. Clementini, P. di Felice, An algebraic model for spatial objects with undetermined boundaries, in: P.A. Burrough, A. Frank (Eds.), Geographic Objects with Indeterminate Boundaries, Taylor & Francis, London, 1996, pp. 155–169.

[17] A.G. Cohn, N.M. Gotts, The 'egg-yolk' representation of regions with undetermined boundaries, in: P.A. Burrough, A. Frank (Eds.), Geographic Objects with Indeterminate Boundaries, Taylor & Francis, London, 1996, pp. 171–187.

[18] M.B. Dale, Some fuzzy approaches to phytosociology: ideals and instances, Folia Geobot. Phytotaxon. 23 (1988) 239–274.

[19] D.A. Davidson, S.P. Theocharopoulos, R.J. Bloksma, A land evaluation project in Greece using GIS and based on Boolean fuzzy set methodologies, Internat. J. Geograph. Inform. Systems 8 (1994) 369–384.

[20] D. Dubois, H. Prade, Fuzzy sets: a convenient fiction for modeling vagueness and possibility, IEEE Trans. Fuzzy Systems 2 (1994) 16–21.

[21] S.M. Engel, With Good Reason: An Introduction to Informal Fallacies, 2nd ed., St Martin's Press, New York, 1982.

[22] G. Evans, Can there be vague objects?, Analysis 38 (1978) 208.

[23] P.F. Fisher, Knowledge-based approaches to determining and correcting areas of unreliability in geographic databases, in: M. Goodchild, S. Gopal (Eds.), Accuracy of Spatial Databases, Taylor & Francis, London, 1989, pp. 45–54.

[24] P.F. Fisher, Probable and fuzzy models of the viewshed operation, in: M. Worboys (Ed.), Innovations in GIS 1, Taylor & Francis, London, 1994, pp. 161–175.

[25] P.F. Fisher, Boolean and fuzzy regions, in: P.A. Burrough, A. Frank (Eds.), Geographic Objects with Indeterminate Boundaries, Taylor & Francis, London, 1996, pp. 87–94.

[26] P.F. Fisher, The pixel: a snare and a delusion, Internat. J. Remote Sensing 18 (1997) 679–685.

[27] P.F. Fisher, T. Orf, An investigation of the meaning of near and close on a university campus, Comput. Environ. Urban Systems 15 (1991) 23–35.

[28] P.F. Fisher, S. Pathirana, The evaluation of fuzzy membership of land cover classes in the suburban zone, Remote Sensing Environ. 34 (1990) 121–132.

[29] P.F. Fisher, S. Pathirana, The ordering of multitemporal fuzzy land-cover information derived from Landsat MSS data, GeoCarto Internat. 8 (3) (1993) 5–14.

[30] P.F. Fisher, J. Wood, What is a Mountain? or The Englishman who went up a Boolean geographical concept but realized it was fuzzy, Geography 83 (1998) 247–256.

[31] G.M. Foody, A fuzzy sets approach to the representation of vegetation continua from remotely sensed data: an example from lowland heath, Photogram. Eng. Remote Sensing 58 (1992) 221–225.

[32] G.M. Foody, Ordinal-level classification of sub-pixel tropical forest cover, Photogram. Eng. Remote Sensing 60 (1994) 61–65.

[33] G.M. Foody, Approaches to the production and evaluation of fuzzy land cover classification from remotely-sensed data, Internat. J. Remote Sensing 17 (1996) 1317–1340.

[34] G.M. Foody, N.A. Campbell, N.M. Trodd, T.F. Wood, Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification, Photogram. Eng. Remote Sensing 58 (1992) 1335–1341.

[35] S. Gale, Inexactness fuzzy sets and the foundation of behavioral geography, Geograph. Anal. 4 (1972) 337–349.

[36] P.T. Geach, The law of excluded middle, Aristotelian Society (Suppl.) 30 (1956) 59–73.

[37] M. Goodchild, S. Gopal (Eds.), Accuracy of Spatial Databases, Taylor & Francis, London, 1989.

[38] S.C. Guptill, J.L. Morrison (Eds.), Elements of Spatial Data Quality, Pergamon, Oxford, 1995.

[39] S. Halldén, The Logic of Nonsense, Uppsala, Universitets Arsskrift, 1949.
[40] D.W. Harvey, Explanation in Geography, Arnold, London, 1969.
[41] A. Kaplan, H.F. Schott, A calculus for empirical classes, Methodos 3 (1951) 165–188.
[42] G. King, Mapping Reality: An Exploration of Cultural Cartographies, Macmillan, London, 1996.
[43] G.J. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice-Hall, Englewood Cliffs, NJ, 1995.
[44] B. Kosko, Fuzziness vs. probability, Internat. J. General Systems 17 (1990) 211–240.
[45] B. Kosko, Fuzzy Thinking: The New Science of Fuzzy Logic, Hyperion, New York, 1993.
[46] R. Kruse, J. Gebhardt, F. Klawonn, Foundations of Fuzzy Systems, Wiley, Chichester, 1994.
[47] P. Lagacherie, P. Andrieux, R. Bouzigues, The soil boundaries: from reality to coding in GIS, in: P.A. Burrough, A. Frank (Eds.), Spatial Conceptual Models for Geographic Objects with Undetermined Boundaries, Taylor & Francis, London, 1996, pp. 275–286.
[48] R. Laurini, D. Thompson, Fundamentals of Spatial Information Systems, Academic Press, London, 1992.
[49] M. Laviolette, J.W. Seaman, The efficacy of fuzzy representations of uncertainty, IEEE Trans. Fuzzy Systems 2 (1994) 4–15.
[50] Y.C. Leung, Locational choice: a fuzzy set approach, Geograph. Bull. 15 (1979) 28–34.
[51] Y.C. Leung, On the imprecision of boundaries, Geograph. Anal. 19 (1987) 125–151.
[52] Y.C. Leung, Spatial Analysis and Planning under Imprecision, Elsevier, New York, 1988.
[53] D.V. Lindley, Comments on The efficacy of fuzzy representations of uncertainty, IEEE Trans. Fuzzy Systems 2 (1994) 37.
[54] J. Lukasiewicz, A. Tarski, Unterschungen über den Aussagenkalkül, C.R. Soc. Sci. Lett. Varsovie Cl 3 23 (1930) 1–21, 30–50.
[55] D. Martin, Mapping population data from zone centroid locations, Trans. Inst. British Geographers NS 14 (1989) 90–97.
[56] D. Martin, Geographic Information Systems: Socioeconomic Applications, Routledge, London, 1996.
[57] D. Martin, I. Bracken, Techniques for modelling population-related raster databases, Environ. Planning A 23 (1991) 1069–1075.
[58] A.B. McBratney, J.J. De Gruijter, A continuum approach to soil classification by modified fuzzy $k$-means with extragrades, J. Soil Sci. 43 (1984) 159–175.
[59] A.B. McBratney, A.W. Moore, Application of fuzzy sets to climatic classification, Agricult. Forest Meteorol. 35 (1985) 165–185.
[60] C. Monger, The Englishman who went up a hill but came down a mountain, Corgi Books, London, 1995.
[61] I.R. Moraczewski, Fuzzy logic for phytosociology 1: syntaxa as vague concepts, Vegetatio 106 (1993) 1–11.
[62] I.R. Moraczewski, Fuzzy logic for phytosociology 2: generalization and prediction, Vegetatio 106 (1993) 13–20.
[63] G. Munda, Multicriteria Evaluation in a Fuzzy Environment: Theory and applications in ecological economics, Physica-Verlag, Heidelberg, 1995.
[64] J.H. Paterson, North America, 9th ed., Oxford University Press, Oxford, 1994.
[65] Z. Pawlak, Rough sets, Internat. J. Comput. Inform. Sci. 11 (1982) 341–356.
[66] J.S. Pipkin, Fuzzy sets and spatial choice, Ann. Assoc. Amer. Geograph. 68 (1978) 196–204.
[67] C. Ponsard, Hierarchie des places centrales et graphes psiflous, Environ. Planning A 9 (1977) 1233–1252.
[68] D.W. Roberts, Ordination on the basis of fuzzy set theory, Vegetatio 66 (1986) 123–131.
[69] D.W. Roberts, Fuzzy systems vegetation theory, Vegetatio 83 (1989) 71–80.
[70] D.W. Roberts, Analysis of forest succession with fuzzy graph theory, Ecol. Modelling 45 (1989b) 261–274.
[71] V.B. Robinson, Some implications of fuzzy set theory applied to geographic databases, Comput. Environ. Urban Systems 12 (1988) 89–98.
[72] V.B. Robinson, Interactive machine acquisition of a fuzzy spatial relation, Comput. Geosci. 16 (6) (1990) 857–872.
[73] V.B. Robinson, A.H. Strahler, Issues in designing geographic information systems under conditions of inexactness, in: Proc. 10th Internat. Symp. on Machine Processing of Remotely Sensed Data, Purdue University, Lafayette, pp. 198–204.
[74] V.B. Robinson, D. Thongs, Fuzzy set theory applied to the mixed pixel problem of multispectral landcover databases, in: B. Opitz (Ed.), Geographic Information Systems in Government, A Deerpak Publishing, Hampton, 1986, pp. 871–885.
[75] B. Russell, Vagueness, Australian J. Philos. 1 (1923) 84–92.
[76] R.M. Sainsbury, What is a vague object?, Analysis 49 (1989) 99–103.
[77] R.M. Sainsbury, Paradoxes, 2nd ed., Cambridge University Press, Cambridge, 1995.
[78] R.M. Sainsbury, T. Williamson, Sorites, in: B. Hale, C. Wright (Eds.), Blackwell Companion to the Philosophy of Language, Blackwell, Oxford, 1995.
[79] H. Symonds, Running High: The First Continuous Traverse of the 303 Mountains of Britain and Ireland, Lochar Publishing, Moffat, 1991.
[80] C. Tilley, Phenomenology of Landscape: Places, Paths and Monuments, Berg, Oxford, 1994.
[81] E.L. Usery, A conceptual framework and fuzzy set implementation for geographic features, in: P.A. Burrough, A. Frank (Eds.), Geographic Objects with Indeterminate Boundaries, Taylor & Francis, London, 1996, pp. 87–94.
[82] F. Wang, Improving remote sensing image analysis through fuzzy information representation, Photogram. Eng. Remote Sensing 56 (1990) 1163–1169.
[83] F. Wang, Towards a natural language user interface: an approach of fuzzy query, Internat. J. Geograph. Inform. Systems 8 (1994) 143–162.
[84] F. Wang, G.B. Hall, Fuzzy representation of geographical boundaries, Internat. J. Geograph. Inform. Systems 10 (1996) 573–590.

[85] F. Wang, G.B. Hall, Subaryono, Fuzzy information representation and processing in conventional GIS software: database design and application, Internat. J. Geograph. Inform. Systems 4 (1990) 261–283.
[86] G.G. Wilkinson, F. Fierens, I. Kanellopoulos, Integration of neural and statistical approaches in spatial data classification, Geograph. Systems 2 (1995) 1–20.
[87] T. Williamson, Vagueness, Routledge, London, 1994.
[88] L. Wittgenstein, Remarks on Colour, Blackwell, Oxford, 1977.
[89] J. Wood, Scale-based characterisation of digital elevation models, in: D. Parker (Ed.), Innovations in GIS 3, Taylor & Francis, London, 1996, pp. 163–175.
[90] M.F. Worboys, GIS: A Computing Perspective, Taylor & Francis, London, 1995.
[91] M.H. Yeates, B.J. Garner, The North American City, Harper & Row, New York, 1971.
[92] L.A. Zadeh, Fuzzy sets, Inform. and Control 8 (1965) 338–353.
[93] L.A. Zadeh, Fuzzy sets versus probability, Proc. IEEE 68 (1980) 421.