

THE VALIDITY AND USEFULNESS OF LAWS IN GEOGRAPHIC INFORMATION SCIENCE AND GEOGRAPHY

Michael F. Goodchild, National Center for Geographic Information and Analysis, and Department of Geography, University of California, Santa Barbara, CA 93106-4060. Phone 805 893 8049, FAX 805 893 3146, Email good@geog.ucsb.edu

1. INTRODUCTION

My contribution to the session at the New Orleans meetings of the Association of American Geographers at which the ideas that appear in this forum were initially aired was as discussant, following the three main presentations by Barnes, Miller, and Smith. This paper expands on the comments I made during the session, and include reactions to the subsequent contribution of Phillips. Specifically, I am concerned with the validity of Tobler's First Law (TFL), and with its usefulness, both considered in the specific context of geographic information science (GIScience). I give my perspective on the relationship between GIScience and geography below. The paper begins with a brief introduction to GIScience, and the role of general principles in the field. I then discuss the possibility of additional laws, a theme that Phillips has already addressed briefly, and the more general question of whether it is possible to build on Tobler's contribution, and to what effect.

2. LAWS AND GISCIENCE

Geographic information science (GIScience) can be defined as that branch of information science that deals with places on or near the surface of the Earth; or alternatively as the set of fundamental principles underlying the design, testing, and use of geographic information technologies (for alternative definitions of GIScience see Goodchild, 1992a; Wright, Goodchild, and Proctor, 1997; Mark, 2003; and the University Consortium for Geographic Information Science, <http://www.ucgis.org>). The two definitions address essentially distinct aspects of GIScience: its scientific role as a focus of theoretical and empirical knowledge; and its pragmatic role as a source of the principles that can guide the development of technology. This duality is of course found across the sciences, since all fields of scientific activity serve both to simplify the world around us through the identification of general principles, in response to a generally felt human emotional need; and to provide the basis for economic and other activities. Simplification is enshrined in Occam's Razor, and is also a legitimate pedagogic strategy; by focusing on principles, students learn things that are generally true both in space and time, and principles therefore form the foundation of most learning in science. Many scientists can recall the satisfaction that came on first being exposed to the simple beauty of the Bohr atom, and its explanation for the complex chemical behavior of the elements and for the Periodic Table. Many geographers must have felt a similar satisfaction in seeing the beauty of the fit between the Atlantic coasts of South America and Africa, and hearing for the first time about the theory of continental drift.

Judging by statements in other pieces in this forum, there is little general agreement about what makes a principle a law. Smith takes the view that a law must be true *entirely*, and

that a single counterexample is sufficient to refute, while Harvey argues that a law need not be deterministic. If there is a meta-law that all laws be deterministic it is clearly not itself a deterministic law – the existence of the Second Law of Thermodynamics, a stochastic law and the victim of Maxwell’s famous demon, would be sufficient to refute the meta-law, and Newton’s Laws of Motion have retained their titles despite their evident falsification in the relativistic and quantum domains.

Expectations about laws clearly vary across the sciences. It is in principle possible for a human individual to violate any deterministic law about individual behavior, which would appear to deny any possibility of such laws in the social sciences, a theme discussed by Barnes. One is reminded of the experimental physicist Ernest Rutherford’s perhaps apocryphal but much-quoted statement about the social sciences, “the only result that can possibly be obtained is that some do, and some don’t”. If laws need not be deterministic, then surely this one has some claim as the First Law of Social Science. Economists refer to the Law of Utility Maximization, a tautology if utility is inferred from behavior, and at best a norm or a standard, with only a weak claim to empirical validity. But from a more pragmatic perspective the distinction between laws, principles, theories, and norms may be moot: if a general statement is empirically valid, its value results from its simplicity, its elegance, and its utility in design and prediction, not from its title. The terms seem virtually interchangeable in common usage, despite the best efforts of philosophers of science to impose some degree of order.

GIScience has a complex relationship to the discipline of geography. Its focus on information, on the description of the Earth’s surface and near-surface, inevitably links it to the study of form -- of how the world *looks*, rather than of how the world *works*. Taylor (1991) made much of this distinction in his critique of GIS as overly concerned with facts, and thus of marginal value in advancing geographic knowledge (but see the response by Goodchild, 1991). But facts provide the context for geographic processes, the evidence from which processes are inferred, and the boundary conditions needed to simulate their effects. Moreover, geographic information systems (GIS) can be seen as tools to take geographic knowledge, conceptualized as theories, expressed as models, and implemented as algorithms, and to examine the implications of such knowledge in the specific geographic settings represented by the GIS’s database of facts.

As such, GIS help to resolve the timeless dilemma in geography about whether to focus on the local or the global – whether geographic science should be idiographic or nomothetic. As Phillips suggests in his discussion of L and G, there is increasing sympathy in many disciplines, including geography, for a middle position in which the specific details of law-like statements are allowed to vary geographically. Recent contributions to the tools of spatial analysis, such as Geographically Weighted Regression (Fotheringham, Brunson, and Charlton, 2002), now provide the techniques to implement this interesting methodological position.

These arguments provide the context to a discussion of the validity and value of TFL in GIScience. It can be read as a statement either about form (interpreting *related* in the sense of *similar*), or about process (interpreting *related* mechanistically), but this

discussion focuses much more on the former. Interest in TFL has grown rapidly in the past decade, largely in tandem with growth in GIScience. As GIS has become more popular and widely adopted, more and more attention has been paid to its fundamentals, and there has been a resurgence of interest in such topics as cartography (and its generalization in GIS visualization), quantitative geography, and spatial statistics. GIScientists have begun to ransack the research record of the 1950s through the 1970s, in a search for methods and theories of particular relevance to GIS design and application. The prominence now attached to TFL in the GIScience literature (see, for example, Longley *et al.*, 2001) is a direct result of this new activity.

3. THE VALUE OF TFL IN GISCIENCE

The construction of digital representations often requires us to recognize and make explicit, and often to formalize tendencies and truths that are so obvious and so ubiquitous that they are rarely noted in everyday life. Consider spatial interpolation, the process by which systems make estimates of the values of variables at places where they have not been measured, based on measurements at other places. Every weather map relies on this process, taking point observations of such variables as temperature or pressure from weather stations, and interpolating between them to obtain complete surfaces, commonly expressed in the form of isolines. No one would deny the logic of estimating the temperature at a point from the nearest available measurements, yet it is in effect an implementation of TFL, and relies implicitly on the law's validity and generality; if TFL were not true, the process would create evident nonsense. A similar process allows estimates to be made of the total ore to be extracted from mines, or the total water resources in a groundwater basin. When necessary, it is possible to express the principle formally, as a measure of spatial autocorrelation or covariance, or as a variogram or covariogram function. Indeed, TFL forms the conceptual foundation for the entire field of geostatistics (Burrough and MacDonnell, 1998; Cressie, 1991; Goovaerts, 1997), and its cousins in other disciplines.

Perhaps the easiest way to see the validity and value of TFL is through a thought experiment, by trying to imagine a world in which it is not true. In such a world, the full range of conditions could be encountered in every minute portion of the world. Every room, for example, might contain the full observed range of the Earth's topographic variation, from the bottom of the Marianas Trench to the summit of Mount Everest; and the full range of climatic conditions as well. This is the world of white noise, a world that would be impossible to describe, live in, sense, navigate through, or farm. All of our methods for simplifying and describing the geographic world, and working and moving within it, rely on the validity of TFL.

TFL is thus obvious in hindsight, and surely no more than common sense. But so too for that matter are Newton's Laws of Motion, or the Second Law of Thermodynamics. Clearly it is the act of discovery of a scientific law that is important in the history of science, together with the context in which the discovery was made and the increment to knowledge that the discovery provided, rather than the law itself. TFL is also close to everyday human experience, a characteristic that it shares in common with geography

and the disciplines of social science, and that denies it the kind of exotic attraction that astronomers and polar scientists can assume for their work. But while arguments like these might explain the lack of attention to TFL, relative to other laws in science, they cannot detract from the law's value and importance in practice.

TFL is stated in a charmingly informal way, and one wonders if it would be taken more seriously if it were formalized, and made a little less accessible. It might state, for example, that for every geographic variable (a function of location $z = f(\mathbf{x})$) there exists some distance d below which covariance is monotonically increasing – or that there exists at least one scale for which spatial autocorrelation is positive. Formalization might address some of the reluctance to give TFL greater stature; but at the same time it would reduce both its accessibility and its charm.

As a general principle, TFL underlies many aspects of GIS design. It explains why it is possible to characterize complex geographic landscapes using a high degree of generalization, with large polygons representing essentially homogeneous regions. It explains why it is possible to represent surfaces using isolines, and why such isolines are not infinitely long and contorted. It explains why spatial interpolation, resampling, contour mapping, and all of the other procedures that involve inferring complete surfaces from a finite set of distributed measurements succeed in providing reasonably useful estimates. It explains why it is possible to rely on GIS databases to provide accurate estimates of relative location, or relative elevation, despite poor levels of absolute accuracy (Zhang and Goodchild, 2002). And it explains the value of such structures as the quadtree, and indeed all of the important GIS data models.

4. BEYOND TFL

If one accepts the First Law as valid and useful, then are there prospects of additional laws? A possible candidate for the Second Law is the principle of spatial heterogeneity discussed by Harvey. Spatial heterogeneity, or non-stationarity in the statistical meaning of that term, implies that geographic variables exhibit uncontrolled variance. There is no concept of an average place on the Earth's surface, comparable for example to the concept of an average human. In the study of time series that exhibit uncontrolled variance there is a concept known as the Noah Effect, which essentially states that an extreme event of any magnitude is possible, if one is willing to wait long enough. One might formulate an equivalent spatial concept, perhaps the Eden Effect, which might state that any extreme condition is possible on the Earth's surface, if one searches far enough. Of course the Earth's surface has limited extent, unlike time, so the parallel is not exact. A corollary of uncontrolled variance in space is that the results of any analysis depend explicitly on the bounds of the analysis: move the study area, and the results will change. Since spatial heterogeneity is a first-order effect, concerning places taken one at a time, and spatial dependence, the basis of TFL, is a second-order effect of places compared two at a time it might make more sense if TFL were the second, rather than the first law.

It is easy to advance other candidates for the status of laws of geography. There is the fractal principle: that geographic phenomena reveal more detail the more closely one

looks; and that this process reveals additional detail at an orderly and predictable rate (Goodchild and Mark, 1987; Mandelbrot, 1982). There is the principle that two distinct conceptualizations of geographic information are possible, as collections of countable, discrete objects littering an otherwise empty space, and as a finite set of continuous fields, or functions of location (Goodchild, 1992b; Worboys, 1995). There is the uncertainty principle, that the geographic world is infinitely complex, and that any representation must therefore contain elements of uncertainty; that many definitions used in acquiring geographic data contain elements of vagueness; and that it is impossible to measure location on the Earth's surface exactly (Zhang and Goodchild, 2002). All of these and many more constitute the foundation principles of GIScience, and merit consideration as potential laws of geography.

Another way to extend TFL is to ask whether it is also true of other spaces. The geographic world is a particular space, a *geospace* in Miller's terms, defined by its extent, its geometry, and its range of useful resolutions. There are many other spaces, such as the space of the cosmos, of the human body, of the human brain, or of the genome, so one might reasonably ask whether TFL is true of all of them. Clearly it is not true of the space of the digits of π , which display as much variation over short distances as over long distances. It would be interesting to determine whether and how TFL applies to the spaces being created by the Internet (Dodge and Kitchin, 2001), and the *small-world* (Watts, 2003) spaces defined by social rather than spatial interaction.

4. DISCUSSION

In retrospect it is clear that we have not yet built effectively on the foundation provided over 30 years ago by TFL. Very little consensus exists on the principles underlying GIS – the principles of GIScience – and therefore on the core content of a GIScience course. Longley *et al.* (2001) include a chapter on the nature of geographic data in their introductory text, but in general it is as hard to find clear statements of the founding principles of the emerging discipline of GIScience as it is to find clear statements of the founding principles of geography. What do we know to be always true of geographic data, and which of those principles apply to what other spaces? The answers to questions like these can serve many useful purposes, from teaching to GIS design, and can address the interesting question of the applicability of GIS techniques to other spaces, an area that is only now being explored (see, for example, <https://www.vbi.vt.edu/article/articleview/48/1/15/>).

From this somewhat utilitarian perspective it is difficult to see why there is such resistance to the idea of laws in GIScience, or in geography. I must admit to being mystified by Barnes' insistence on "unsettling" TFL, and unable to follow his argument about local laws, and the example of Boyles' Law. I remember high-school instructors in science who at the end of a term would attempt to lighten the otherwise serious atmosphere of the class by devoting the last session to the failed theories of science – phlogiston, for example – and to the life histories of famous scientists, content that is certainly interesting at an anecdotal level, and at some level may help people understand

a law's contribution to science. But we cannot escape the fact that Boyle's Law is universal, true, and eminently useful.

I joined geography as a graduate student, after an undergraduate career in physics, and I remember being struck in my first course in urban geography, taught with great passion by Gerard Rushton, by the empirical status of Christaller's Central Place Theory. Here is a theory that clearly is not in agreement with empirical reality, and yet unlike phlogiston it remains a core part of the geography curriculum. It took me some time to realize that Central Place Theory, the Law of Utility Maximization, and many other principles in the social sciences remain important because they provide norms against which empirical reality can be compared, and simplify our thinking about systems that might otherwise be incomprehensible, ideas that have few parallels in physics. But TFL seems rather different, and more in the style of the traditional laws of science, which are judged primarily for their empirical validity, and obtain their value from their simplicity and their practical usefulness.

REFERENCES

Burrough, P.A., and R.A. MacDonnell, 1998. *Principles of Geographical Information Systems*. New York: Oxford University Press.

Cressie, N.A.C., 1991. *Statistics for Spatial Data*. New York: Wiley.

Dodge, M., and R. Kitchin, 2001. *Mapping Cyberspace*. New York: Routledge.

Fotheringham, A.S., C. Brunsdon, and M. Charlton, 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester, UK: Wiley.

Goodchild, M.F., 1991. Just the facts. *Political Geography Quarterly* 10: 192-193.

Goodchild, M.F., 1992a. Geographical information science. *International Journal of Geographical Information Systems* 6(1): 31-45.

Goodchild, M.F., 1992b. Geographic data modeling. *Computers and Geosciences* 18(4): 401-408.

Goodchild, M.F., and D.M. Mark, 1982. The fractal nature of geographic phenomena. *Annals of the Association of American Geographers* 77: 265-278.

Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press.

Longley, P.A., M.F. Goodchild, D.J. Maguire, and D.W. Rhind, 2001. *Geographical Information Systems and Science*. New York: Wiley.

Mandelbrot, B.B., 1982. *The Fractal Geometry of Nature*. San Francisco: Freeman.

Mark, D.M., 2003. Geographic information science: defining the field. In M. Duckham, M.F. Goodchild, and M.F. Worboys, editors, *Foundations of Geographic Information Science*. New York: Taylor and Francis, pp. 3-18.

Taylor, P.J., 1991. GKS. *Political Science Quarterly* 9: 211-212.

Watts, D.J., 2003. *Six Degrees: The Science of a Connected Age*. New York: Morton.

Worboys, M.F., 1995. *GIS: A Computing Perspective*. New York: Taylor and Francis.

Wright, D.J., M.F. Goodchild, and J.D. Proctor, 1997. Demystifying the persistent ambiguity of GIS as 'tool' versus 'science'. *Annals of the Association of American Geographers* 87(2): 346-362.

Zhang, J.X., and M.F. Goodchild, 2002. *Uncertainty in Geographical Information*. New York: Taylor and Francis.