FORUM

Tobler's First Law and Spatial Analysis

Harvey J. Miller

Department of Geography, University of Utah

I never thought that others would take them so much more seriously then I did.

—Albert Einstein on his theories

"I invoke the first law of geography: everything is related to everything else, but near things are more related than distant things" (Tobler 1970). How could a sentence justifying heuristic calculations in a crude urban growth simulation generate an icon now known as Tobler's First Law (TFL)? Why has this law resonated so strongly in geography?

Waldo Tobler could invoke a first law of geography since the proposition that near things are more related seemed reasonable in 1970. It is enduring since near and related are useful concepts at the core of spatial analysis and modeling. And in 2004 and beyond, TFL is still useful since the rise of geographic information science and technologies allow greater sophistication when measuring and analyzing these concepts. This is ironic considering that Tobler apparently invoked the law in part to apologize for the slow computers at that time.

I am going to sidestep the issue of whether TFL is in fact a law by noting that science accepts the concept of *empirical laws*, or compact descriptions of patterns and regularities. These are not required to be immutable truths (Casti 1990; Swartz 2001). We certainly have ample evidence to support TFL: you may have noticed on the way to work this morning that the world is orderly with respect to space. Scientific laws are also not required to be causal, for example, Newton's Law of Gravity is not an explanation. Although not causal, TFL is consistent with an elegant process argument: overcoming space requires expenditure of energy and resources, something that nature and humans try to minimize (although not exclusively, of course). I accept TFL as reasonable regularity that generally holds true.

The issues I am going to examine are the central roles of "near" and "related" to spatial analysis and the increasing levels of sophistication that we can achieve when measuring and analyzing these concepts. I also suggest that relations among near entities do not imply a simple, sterile geography; complex geographic processes and structures can emerge from local interactions. Indeed, the sensitivity of geographic and other phenomena to local interactions implies that we should carefully measure and analyze relations among near things.

What Is Related?

What do we mean when we say that two geographic entities are related? At the very least, we are claiming that there is a positive or negative correlation between these entities. Spatial association does not necessarily imply causality. Two things that are associated may be involved in a causal relationship, or there may be other hidden variables that cause the association. Although correlation is not causality, it provides evidence of causality that can (and should) be assessed in light of theory and/or other evidence.

TFL is at the core of spatial autocorrelation statistics, that is, quantitative techniques for analyzing correlation relative to distance or connectivity relationships. Although spatial autocorrelation is often treated as confounding (e.g., something to be corrected in regression modeling), it is information bearing since it reveals the spatial associations among geographic entities. In 1970, techniques for measuring and analyzing spatial autocorrelation were crude, providing only a single, summary number for an entire spatial dataset indicating the overall intensity of the spatial association. Spatial analysts now recognize every location has an intrinsic degree of uniqueness due to its situation relative to the rest of the spatial system. Similar to spatial autocorrelation, *spatial heterogeneity* is not just parameter drift to be corrected: it is information bearing since it reveals both the intensity and pattern of spatial associations. Disaggregate spatial statistics such as local indicators of spatial association (LISA) statistics (Anselin 1995), the G statistics (Getis and Ord 1992) and geographically weighted regression (Brunsdon, Fotheringham, and Charlton 1996) capture spatial association and heterogeneity simultaneously.

Annals of the Association of American Geographers, 94(2), 2004, pp. 284–289 © 2004 by Association of American Geographers
Initial submission, July 2003; final acceptance, November 2003
Published by Blackwell Publishing, 350 Main Street, Malden, MA 02148, and 9600 Garsington Road, Oxford OX4 2DQ, U.K.

These techniques generate abundant information that can be used in both exploratory and confirmatory analysis to generate and test hypotheses about spatial relations. Their data requirements and demands on geo-visualization techniques make them unimaginable prior to the rise of widely available digital geo-data and GIS.

Another core spatial analytic technique that exploits TFL is *spatial interpolation* or techniques for generating missing or hidden variables in geographic space. Some of these techniques are very sophisticated in their implementation of TFL. For example, *kriging* treats the spatial variable being interpolated as *regionalized*, meaning that it varies continuously across space according to some spatial lag or distance in a partly random and partly deterministic manner. This admits a wide range of distance functions and clustering patterns. It also allows ad-hoc adjustments based on qualitative information. Despite this flexibility, kriging is also powerful in the sense that there are well-established techniques for estimating parameters that minimize interpolation error, given sample data and a hypothesized spatial lag model. These error measures are spatially disaggregate and can be mapped and visualized, providing a detailed record of interpolation accuracy across space (see Lam 1983; Isaaks and Srivastava 1989; Oliver and Webster 1990).

A stricter type of spatial association is *spatial interaction*, or the movement of individuals, material, or information between two geographic locations. Spatial interaction is closely related to spatial autocorrelation: spatial interaction models are special cases of a general model of spatial autocorrelation (Getis 1991). Similar to spatial autocorrelation, advanced techniques for spatial interaction and spatial choice modeling recognize spatial heterogeneity or map pattern effects. These effects result from individuals simplifying spatial choice problems by clustering or lumping choices together, often based on proximity (Fotheringham 1983; Kanaroglou and Ferguson 1996; Bhat, Govindarajan, and Pulugurta 1998). Computational techniques, such as genetic algorithm-based parameter estimation and artificial neural networks, are improving the robustness of spatial interaction modeling for noisy and nonquantitative data (Dougherty 1995; Diplock and Openshaw 1996).

What Is Near?

The discussion in the previous section leaves the concept of near as vague and undefined as Waldo Tobler did when invoking TFL. This section, based on Miller and Wentz (2003), suggests that near is central to spatial analysis. It also suggests that near is a more flexible and powerful concept than commonly appreciated.

As Gatrell (1983) points out in his excellent book *Distance and Space*, geographers do not have a solitary claim on the concept of space; we can form a mathematical space by defining a set of objects and relations between all pairings of these objects. These relations can be quantitative or qualitative. However, as geographers, we are really only interested in a subset of all possible spaces, namely, *geo-spaces* or those that can be meaningfully represent phenomena on or near the surface of the Earth.

What distinguishes geo-spaces from other spaces? In geo-spaces, the objects correspond to locations on the surface of the Earth (at least conceptually) with defined *shortest path relations* between all pairings. These are the minimum-cost routes for physical movement or virtual interaction between objects, where cost is interpreted generally. The shortest-path relations determine the measurement and analysis of geographic attributes (Beguin and Thisee 1979).

In most of the geographic and related literature, nearness is typically defined based on the straight-line segment connecting two locations, that is, the Euclidean distance for the location pair. This is only one possibility. There are an infinite number of shortest-path relations that obey the *metric space* conditions of symmetry, nonnegativity, and triangular inequality (Love, Morris, and Wesolowsky 1988; Puu and Beckmann 1999). If we are willing to relax these metric requirements so that only the triangular inequality condition holds, the resulting space is a *quasi-metric*. This can still support measurement and spatial analysis (Huriot, Smith, and Thisse 1989; Smith 1989).

Geographic phenomena that do not appear to be consistent with TFL may, in fact, be following non-Euclidean nearness relations. This can include geographic diffusion processes such as disease propagation (Cliff and Haggett 1998), movement and interaction at the urban, regional, and national scales (Worboys, Mason, and Lingham 1998; Puu and Beckmann 1999) and human perception of geographic space (Montello 1992). Waldo Tobler has spent much of his career trying to convince us that non-Euclidean geo-spaces are also meaningful using cartographic transformations and other clever analytical and visualization techniques (e.g., Tobler 1976a, 1976b, 1978, 1987, 1994).

Nearness relations need not be restricted to empty space. Some geographic phenomena are conditioned by geographic attributes such as terrain, land cover, and traffic congestion. To capture these effects, we can generalize the concept of distance to least-cost paths through geographic space (Angel and Hyman 1976). This requires treating a spatially continuous attribute or attributes as a cost field that affects movement or interaction. This is a well-studied problem in spatial analysis and geographic information science; several tractable computational algorithms are available for special cases of this general problem (e.g., Smith, Peng, and Gahinet 1989; de Berg and van Kreveld 1997).

Nearness is a central organizing principle of geo-space, but it is not required to be a function of Euclidean, metric, or even an empty space. There are a wide range of analytical and computational techniques for representing and analyzing these spaces and no reason in principle why they should not be part of a standard GIS toolkit.

But Isn't the World Shrinking?

Distance was meaningful when von Thünen contemplated the ponderous movements of oxcarts between his farm and a central market. The past two centuries have witnessed *space-time convergence*: transportation and communication technologies have shrunk the world to an incredible degree. Locations on the Earth's surface are much closer to each other with respect to the time required for movement and interaction (Janelle 1969). Does this make TFL trivial, since many things are now near?

Waldo Tobler addressed this issue in a 1999 address to the ESRI User Conference. Tobler noted that while the world is shrinking, it is also *shriveling*; relative differences in transportation and communication costs are increasing at most geographic scales. When transportation technology was limited to biological or wind power, all persons, whether noble or peasant, could move only at the same slow speed, albeit with different levels of comfort. The automobile and airplane make the world much smaller, but only if these technologies are accessible and affordable for you. As population growth and urbanization continue, some transportation networks are becoming saturated and congested, creating a complex geography of accessibility associated with differing abilities to pay the housing costs required to avoid long commute times. Transportation cost differentials across space increase when networks (such as airlines and railroads) are pruned and concentrated for economic efficiency or when cities or regions experience collapse of their transportation infrastructure (examples include sub-Saharan Africa and Afghanistan).

Couclelis and Getis (2000) note that the world is also *fragmenting*: many activities are becoming more loosely connected to geographic space. With portable computing and communications technologies such as laptops and cell phones, a person can work at the office, at home, at a coffee shop, or in a park. Thus, there is no longer a single unequivocal location that can be associated with work activities. However, this is predicated on the availability and affordability of telecommunications technologies, and these are still out of reach for many individuals and families in the United States and elsewhere in the world.

Is TFL still valid in a shrinking and fragmenting world? The question is whether near and distant are still valid concepts in this world. The "Death of Distance" argument that dominated much of the early literature on the Internet and cyberspace (e.g., Caincross 1997) is simplistic because it assumes that communication has only a substitution relationship with transportation (i.e., more virtual interaction implies less physical movement). In fact, empirical evidence suggests that the opposite is the case: the rise of telecommunication demand has been paralleled by a corresponding increase in travel demand at all geographic scales (Couclelis 2000). Many of the central places at the end of the Industrial Age are still central in the Information Age. Locations such as Midtown Manhattan and Soho-London are still highly desirable for corporate headquarters, particularly for supposedly footloose activities such as decision making and creative work (Graham and Marvin 1996). When people and corporations have more freedom over where to locate, many chose to locate even closer to each other.

Nearness as a concept can be extended to include both space and time. Janelle (1995) classifies communication modes based on their spatial and temporal constraints. Spatial constraints require either physical presence or telepresence, while temporal constraints require either synchronous or asynchronous activity. This leads to four communication modes as indicated in Table 1. *Synchronous presence* (SP) is the time-honored communication mode of face-to-face (F2F) interaction. F2F requires coincidence in both time and space. *Synchronous telepresence* (ST) requires only coincidence in time: telephones, radio, and TV allow individuals to communicate among different places at the same time. *Asynchronous presence* (AP) requires coincidence in space but not time: examples include Post-It® notes and hospital charts. *Asynchronous telepresence* (AT) does not require coincidence in space and time. Printed media, e-mail, and Web pages are popular examples of AT.

Figure 1 characterizes the communication modes in Table 1 in space and time. The thick arrows indicate two persons' movements in space with respect to time, while

Table 1. Spatial and Temporal Constraints on Communications (based on Harvey and Macnab 2000; Janelle 1995)

	Spatial	
Temporal	Physical Presence	Telepresence
Synchronous	**SP** Face to face (F2F)	**ST** Telephone Instant messaging Television Radio Teleconferencing
Asynchronous	**AP** Refrigerator notes Hospital charts	**AT** Mail Email Fax machines Printed media Web pages

thin lines indicate communication. Communication can only occur at specific locations or *space-time stations* that allow this activity. Not shown is the possibility of mobile communication, that is, a space-time communication station that follows a person's movement in space (e.g., mobile phones, wireless Internet clients).

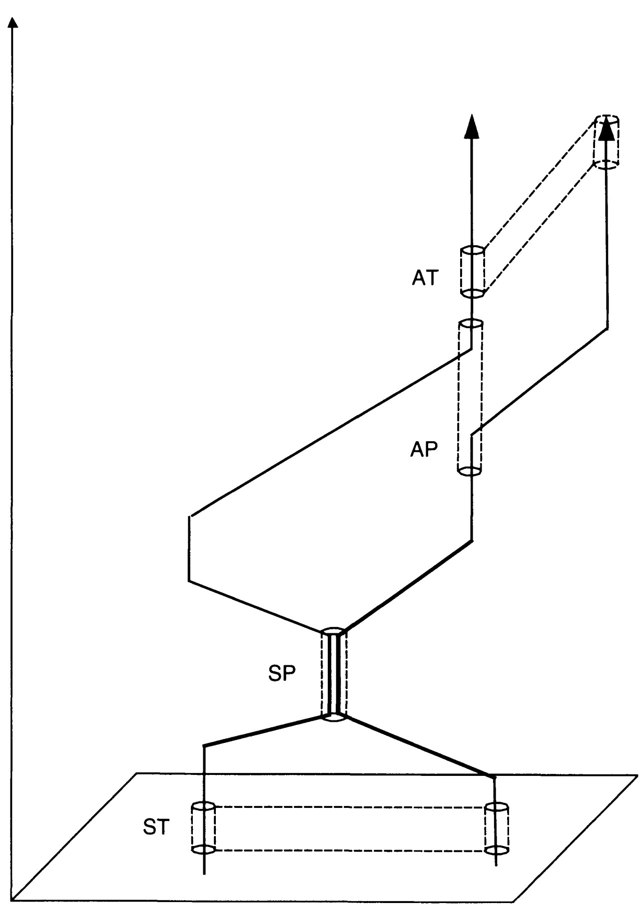

Figure 1. Presence and telepresence in space-time paths.

This discussion suggests, at least in a highly preliminary manner, that the concept of near in geography could be expanded to include both space and time. As Harvey and Macnab (2000) argue, while the role of space may be diminishing for some types of communication, temporal coincidence remains a prerequisite. In the realm of real-time communications the fundamental geographic concept of region may need a temporal overhaul. In some cases, it is not just a matter of where you are, but also *when* you are.

In recent years, space-time analysis has experienced a renaissance as researchers, encouraged by developments in GIS, have expanded their power and scope of theories and techniques for analyzing space-time behavior at multiple scales. The rapidly improving ability to collect space-time activity data through information technologies such as cellular phones, wireless PDAs, global positioning system receivers, and radiolocation methods is improving the quantity and quality of these data and reducing their cost. There are also parallel developments in geographic information science such as spatio-temporal databases, multidimensional GIS, geographic data mining, and geographic visualization (Miller 2003).

Near Is Beautiful

The importance that TFL places on things that are near may be criticized as a simplistic view of geography. Surely the world is more complex than can be explained by simple relations among things that are near! Can we explain geographic phenomena as elaborate as an ecosystem or an economy by only looking at spatial relations among local things?

Complex adaptive systems (CAS) theory suggests that near can be sufficient: simple, local interactions among entities can produce complex global behavior that is not completely predictable or controllable. "Emerge" is a precise term; it means that global behavior is not evident directly from the small set of rules that describe each individual's behavior. Complexity literally emerges from the interactions of simple behaviors (Manson 2001). Local interactions are capable of generating complex aggregate dynamics and intricate structures in space and time (Flake 1998).

The rise of CAS and other complexity theories over the last three decades suggests that near is a valid concept for understanding many real-world phenomena. Relations among things that are near can generate complex spatiotemporal phenomena. Complexity theory also suggests the importance of geographic context: a system's growth and development is sensitive to the pattern and intensity of local interactions. There are insights to be gained by carefully considering, measuring, and analyzing what we mean by near (and distant) in geographical analysis. This is not to suggest that only near things are important, but rather that near is a meaningful starting point for geographic investigation. Indeed, the "small world" phenomena in social and other networks suggests that near and distant can interact to create extensive interconnections in spatial systems (Watts 1999).

Conclusion

My basic argument in this essay is that TFL is a useful law for guiding geographic research, both historically and into the future. TFL is central to core spatial analytical techniques as well as analytical conceptions of geographic space. Continuing progress in spatial analysis as well as the rise of digital geographic databases, geographic information technologies, and geographic information science is breathing new life into TFL. We can measure spatial relations at disaggregate levels, highlighting rather than masking individual-level differences and the role of spatial heterogeneity and spatial context. We can analyze distance and spatial relations using alternative spatial metrics, attributed geographic space, and time. CAS and other computational theories suggest that simple and near are sufficient to generate complex behavior and structures. Indeed, many geographic phenomena may be highly sensitive to relations among near things.

It seems that Waldo Tobler was not completely serious in invoking TFL; the phrase in his 1970 paper reads more like a droll apology than a sober tenet in a *Principia Geographica.* Nevertheless, Tobler articulated a precept that many geographers (and others) continue to find useful and powerful. To a large degree, TFL distinguishes geography from other fields of inquiry: it says that geo-space matters. Using TFL as a core principle, spatial analysis and geographic information science continue to develop sophisticated techniques for extracting explanatory and predictive power from geo-spaces.

Acknowledgments

Thanks to Mike Goodchild, Dan Sui, Tom Cova, and an anonymous referee for their helpful comments. Additional thanks to Dan Sui for organizing the panel discussion and forum.

References

Angel, S., and G. M. Hyman. 1976. *Urban fields: A geometry of movement for regional science*. London: Pion.
Anselin, L. 1995. Local indicators of spatial association–LISA. *Geographical Analysis* 27:93–115.
Beguin, H., and J.-F. Thisse. 1979. An axiomatic approach to geographical space. *Geographical Analysis* 11:325–41.
Bhat, C. R., A. Govindarajan, and V. Pulugurta. 1998. Disaggregate attraction-end choice modeling. *Transportation Research Record* 1645:60–68.
Brunsdon, C., A. S. Fotheringham, and M. E. Charlton. 1996. Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis* 28:281–98.
Caincross, F. 1997. *The death of distance*. Boston: Harvard Business School Press.
Casti, J. L. 1990. *Searching for certainty: What scientists can know about the future*. New York: William Morrow.
Cliff, A. D., and P. Haggett. 1998. On complex geographic space: Computing frameworks for spatial diffusion processes. In *Geocomputation: A primer*, ed. P. A. Longley, S. M. Brooks, R. McDonnell, and B. MacMillan, 231–56. New York: John Wiley and Sons.
Couclelis, H. 2000. From sustainable transportation to sustainable accessibility: Can we afford a new tragedy of the commons? In *Information, place and cyberspace: Issues in accessibility*, ed. D. G. Janelle and D. C. Hodge, 341–56. Berlin: Springer.
Couclelis, H., and A. Getis. 2000. Conceptualizing and measuring accessibility within physical and virtual spaces. In *Information, place and cyberspace: Issues in accessibility*, eds. D. G. Janelle and D. C. Hodge, 15–20. Berlin: Springer.
de Berg, M., and M. van Kreveld. 1997. Trekking in the Alps without freezing or getting tired. *Algorithmica* 18:306–23.
Diplock, G., and S. Openshaw. 1996. Using simple genetic algorithms to calibrate spatial interaction models. *Geographical Analysis* 28:262–79.
Dougherty, M. 1995. A review of neural networks applied to transport. *Transportation Research C* 3:247–60.
Flake, G. W. 1998. *The computational beauty of nature*. Cambridge, MA: MIT Press.
Fotheringham, A. S. 1983. A new set of spatial-interaction models: The theory of competing destinations. *Environment and Planning* A 15:15–36.
Gatrell, A. C. 1983. *Distance and space: A geographical perspective*. Oxford: Clarendon Press.
Getis, A. 1991. Spatial interaction and spatial autocorrelation: A cross-product approach. *Environment and Planning* A 23:1269–77.
Getis, A., and J. K. Ord. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24:189–206.
Graham, S., and S. Marvin. 1996. *Telecommunications and the city: Electronic spaces, urban places*. New York: Routledge.
Harvey, A., and P. A. Macnab. 2000. Who's up? Global interpersonal temporal accessibility. In *Information, place and cyberspace: Issues in accessibility*, ed. D. G. Janelle and D. C. Hodge, 147–70. Berlin: Springer.
Huriot, J-M., T. E. Smith, and J.-F. Thisse. 1989. Minimum-cost distances in spatial analysis. *Geographical Analysis* 21: 294–315.
Isaaks, E. H., and R. M. Srivastava. 1989. *An Introduction to Applied Geostatistics*. New York: Oxford University Press.

Janelle, D. G. 1969. Spatial organization: A model and concept. *Annals of the Association of American Geographers* 59: 348–64.

———. 1995. Metropolitan expansion, telecommuting and transportation. In *The geography of urban transportation*, ed. S. Hanson, 407–34. New York: Guilford.

Kanaroglou, P. S., and M. R. Ferguson. 1996. Discrete spatial choice models for aggregate destinations. *Journal of Regional Science* 36:271–90.

Lam, N. S.-N. 1983. Spatial interpolation methods: A review. *American Cartographer* 10:129–49.

Love, R. F., J. G. Morris, and G. O. Wesolowsky. 1988. *Facility location: Models and methods.* New York: North-Holland.

Manson, S. M. 2001. Simplifying complexity: A review of complexity theory. *Geoforum* 32:405–14.

Miller, H. J. 2003. What about people in geographic information science? *Computers, environment and urban systems* 27: 447–53.

Miller, H. J., and E. Wentz. 2003. Geographic representation and spatial analysis in geographic information systems. *Annals of the Association of American Geographers* 93:574–94.

Montello, D. R. 1992. The geometry of environmental knowledge. In *Theories and methods of spatio-temporal reasoning in geographic space*, ed. A. U. Frank, I. Campari, and U. Formentini, 136–52. Berlin: Springer-Verlag Lecture Notes in Computer Science #639.

Oliver, M. A., and R. Webster. 1990. Kriging: A method for interpolation for geographical information systems. *International Journal of Geographical Information Systems* 4:313–32.

Puu, T., and M. Beckmann. 1999. Continuous space modeling. In *Handbook of transportation science*, ed. R. W. Hall, 269–310. Norwell MA: Kluwer Academic.

Smith, T. E. 1989. Shortest-path distances: An axiomatic approach. *Geographic Analysis* 21:1–31.

Smith, T. R., G. Peng, and P. Gahinet. 1989. Asynchronous, iterative, and parallel procedures for solving the weighted-region least cost path problem. *Geographical Analysis* 21:147–66.

Swartz, N. 2001. Laws of nature. *The Internet dictionary of philosophy*, http://www.utm.edu/research/iep/

Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46: 234–40.

———. 1976a. The geometry of mental maps. In *spatial choice and spatial behavior: Geographic essays on the analysis of preferences and perceptions*, ed. R. G. Golledge and G. Rushton, 69–81. Columbus: Ohio State University Press.

———. 1976b. Spatial interaction patterns. *Journal of Environmental Systems* 6:271–301.

——— 1978. Migration fields. In *Population mobility and residential change*, Northwestern University Studies in Geography Number 25, ed. W. A. V. Clark and E. G. Moore, 215–32. Evanson, IL: Department of Geography, Northwestern University.

———. 1987. Experiments in migration mapping by computer. *American Cartographer* 14:155–63.

———. 1994. Bidimensional regression. *Geographical Analysis* 26:187–212.

Watts, D. J. 1999. *Small worlds.* Princeton, NJ: Princeton University Press.

Worboys, M. F., K. Mason, and J. Lingham. 1998. Computational techniques for non-Euclidean planar spatial data applied to migrant flows. In *Innovations in Geographical Information Systems* 5, ed. C. Carver, 35–45. London: Taylor and Francis.

Correspondence: Department of Geography, University of Utah, 260 S. Central Campus Dr., Room 270, Salt Lake City, UT 84112-9155, e-mail: harvey.miller@geog.utah.edu.