A QUANTITATIVE APPROACH TO INTERESTINGNESS

A. ROSS ECKLER
Morristown, New Jersey

In "Good Grief!" in the November 1973 Word Ways, Merlin X. Houdini IV (a pseudonym of Dmitri Borgmann) stated his Fundamental Axiom of Logology: "All words and all names tantalize with their interest, if you but perceive them correctly". He illustrated this with an analysis of the names of the 15 largest US cities. He repeated this claim under his own name in "The Keystone of Logology" in February 1977, illustrating it with twenty words from the 1974 Merriam-Webster Pocket Dictionary. He even offered a prize of $25 to the first Word Ways reader who could produce a word that he could not show was interesting. There were no takers, probably because it was clear that Borgmann would go to extreme lengths to find an "interesting" property. He demolished my suggestion of *pseudofeverishly,* proposing ten properties such as its 13 different letters (half of the alphabet) and its anagrammability (*fervidly—eh,.spouse*?). In "On the Inter(e)state" in May 1980, Philip Cohen criticized Borgmann on the grounds that the properties he found for the 20 words were shared by many other words, and proposed that one should look for properties that made the word unique within its group. He then proceeded, but with less than full success, to look for unique logological properties of the 50 states. The topic of interestingness languished for eighteen years, before being revived in the November 1998 issue with Dave Morice's "All Numbers Less Than 100 Are Interesting?" in which he looked for logological properties unique with respect to all vigintillion cardinals. Anil sampled an Australian dictionary, reporting in November 2004 "Are All Words Interesting?"..

Interestingness is, of course, a subjective matter—like beauty, in the eye of the beholder. However, just possibly it can be placed in a quantitative framework, the subject of this article. Following Cohen's lead, I propose that (1) any property cited as a proof of interestingness should be simple to state and comprehend, and (2) it should not be shared by any other member of the word set under consideration (or, if this is too stringent, not shared by more than (say) one per cent of the word set).

In abstract terms, consider a set of objects (for example, main entries in a specified dictionary, names of the states, etc.) possessing various properties (for example, word length, numerical score, alphabetic position, number of dictionary-sanctioned transposals, etc). Each property must be specifiable by a number, or at least by a natural ordering (such as the alphabet).

If only one property is considered, only two members of the set (the smallest and the largest) have a claim to interestingness (i.e., uniqueness). But suppose there are two properties under consideration. Now the set can be plotted as a cluster of points on a plane, and more extremals are possible. To fix ideas, think of the members of the set as stakes in the ground. One has a most northerly and a most southerly stake, a most easterly and a most westerly one as well. But there can be other ways of identifying the extremals. Wrap a conceptual string around the set of objects and pull it tight; certain ones of the stakes (in mathematical terms, the convex hull of the set) will touch the string, and each of these are extremal points. How many are there? There can be as few as three, but more likely several more, perhaps six or eight. (It is an interesting problem in mathematics to ascertain the average number of such extremals, given assumptions about the way in which the stakes are scattered on the plane.). With three properties, one ends up not with a taut string but a shrink wrap enclosing the set of objects, with perhaps a dozen or two supporting this covering. (A more general mathematical problem: how fast does the average number of extremals increase as a function of the number of properties?)

One can consider an even more generous definition of extremality. In a set of objects plotted on a plane, there will always exist a subset in which each member of the subset is not simultaneously exceeded in both properties. For example, plot US cities, towns and villages according to their population and their elevation above sea level; there exist certain ones for which no other cities, towns or villages exist that are simultaneously higher and more populous (two such are Denver CO and Leadville CO). Any members of this subset have a legitimate claim to being considered extremal (and therefore interesting).

Internal points in the set may also have a claim to interestingness. Each point in the set has a nearest neighbor, a distance x units away. For which member of the set of objects is x the smallest—that is, which object has the nearest nearest neighbor? (More generally, this can be defined in terms of the nth nearest neighbor.) This seems to be related to the question of what is a coincidence, and is illustrated in "Letters of the Presidents" in the November 2004 Word Ways.

To sum up: a quantitative approach to interestingness consists of (1) limiting the population under consideration to a set of manageable size, and (2) selecting a number of independent simply-stated properties with which to judge the members of the set. If the set is not too large (say ten to fifty), it is likely that all members of the set can be proved interesting, using extremal ideas as outlined above, but for very large sets (such as the numbers from one to vigintillion) it is obvious that a sufficient number of properties cannot be assembled. Nevertheless, Dave Morice succeeded in demonstrating interestingness for the infinitesimal subset 1-99 using a small set of properties!