A NEW APPROACH TO THE TEN-SQUARE

A. ROSS ECKLER
Morristown, New Jersey

Ever since Chris Long's mathematical model in the February 1993 Word Ways showing that it takes a list of at least 250,000 ten-letter words to have a reasonable chance of forming a ten-square, logologists have despaired of forming one from dictionary-sanctioned words. Only by extending the definition of an acceptable "word" does a ten-square seem possible. One way to accomplish this is to allow words from any language (a polyglot square); another is to allow proper or personal names. This article suggests how the latter approach might be realized by reducing the construction of a ten-square to relatively simple independent tasks.

Consider a square formed entirely of five-letter first names and five-letter last names such as ELMER DAVIS or JAMES SMITH. This latter example may be the commonest one in the United States, as documented by http://www.census.gov/genealogy/names, a tabulation of approximately 6.29 million surnames. According to this, Smith forms 1.006 per cent of all surnames, and James 3.318 per cent of all first names. If there are 140,000,000 males in the United States, this suggests the existence of (125,000,000)(.01006)(.03318) = 46731 bearing this name. Such a ten-square can be constructed out of three independent parts: a five-square of first names (male or female), a five-square of surnames, and a double five-square with surnames in one direction and personal names in the other. Steve Root extracted from the above-cited website 1488 five-letter first names and the commonest 1064 five-letter last names, forming from these 3456 first-name squares and 1154 surname squares. Here is a typical first-name square and surname square, each name given with its probability of appearance:

EDGAR	.00080	GRANT	.00060
DIANE	.00359	RILEY	.00053
GAVIN	.00010	ALLEN	.00199
ANITA	.00162	NEESE	.00002
RENAE	.00008	TYNER	.00003

Using a simple scaling argument, one can determine how many first names and surnames are needed to produce, on the average, just one square. (If 2n names yield 32 squares, then n names should, on the average, produce one.) These numbers are 292 and 281 for the first-name and surname squares, only slightly larger than the theoretical value of 250 calculated by Chris Long.

The double five-square with first names horizontally and surnames vertically is much more difficult to find, and in fact Steve Root was able to construct only three out of the same 1488 first names and commonest 1064 last names:

ABDUL	.00007	RHEBA	.00001	RHEBA	.00001
LEORA	.00007	HAZEL	.00161	HAZEL	.00161
GALEN	.00009	OZELL	.00001	OZELL	.00001
ELANE	.00001	NELLA	.00003	NELLE	.00003
RENAY	.00001	ELLEN	.00173	ELLEN	.00173

The corresponding frequencies for the surnames are ALGER .00003, BEALE .00003, DOLAN .00008, URENA .00002, LANEY .00004, RHONE .00002, HAZEL .00004, EZELL .00006, BELLE .00002, ALLAN .00003 and ALLEN .00199. If one takes the square root of the product of the first names (1488) and the surnames (1064) as the effective number of names to apply the scaling argument to, then one double five-square should appear, on the average, if 1104 names are available. Again, this is only slightly larger than Chris Long's theoretical value of 992 for the double five-square, suggesting that one should not be surprised at the small yield.

These squares can be combined to produce a ten-square of possible names. The numbers at the right give the expected number of such individuals among the 140,000,000 males (or females) in the United States, assuming that first names and surnames combine according to the nationwide Census Bureau statistics. Such an assumption is clearly not true for names like ABDUL SMITH, predicted to appear as the name of (140,000,000)(.00007)(.01006) = 99 individuals. (In reality, there are no Social Security beneficiaries, and only three in the AOL White Pages.)

```
E D G A R R H O N E          2.2
D I A N E H A Z E L         20.1
G A V I N E Z E L L          0.6
A N I T A B E L L E          4.5
R E N A E A L L E N         22.3
R H E B A G R A N T          0.8
H A Z E L R I L E Y        119.5
O Z E L L A L L E N          2.8
N E L L E N E E S E          0.1
E L L E N T Y N E R          2.1
```

Purists will object that some of these names are highly unlikely to appear among United States residents—and they are right. A check of the National White Pages on AOL revealed Diane Hazel 3, Anita Belle 3, Renae Allen 12, Hazel Riley 20 and Ozell Allen 9; the 67 million records of Social Security beneficiaries who died prior to November 2001 came up with Edgar Rhone 2, Ozell Allen 2 and Ellen Tyner 3. The remaining three names, Gavin Ezell, Rheba Grant and Nelle Neese, all quite rare according to the Census Bureau statistics, did not appear in a search of various genealogical sites: Rootsweb, Gendex, Family Forum, and the Mormons.

One can, of course, repeat this exercise with names consisting of four-letter first names and six-letter surnames. This leads to much more common first-name squares such as ROSA OPAL SARA ALAN or GLEN LORI ERIN NINA, but six-letter surname squares will probably feature more unusual names than five-letter ones.

We have evaluated only one of the 3(3456)(1154), or nearly 12 million, possible ten-squares; perhaps one of the others has ten "real" names in it. However, it would be very tedious to determine this, even with the aid of a computer that can compare the names against a master list such as the AOL White Pages. I suspect it would be quicker to construct a ten-square directly from such a list; once one has amassed some 300,000 names such a square should certainly appear. Such a ten-square would, of course, include any name totaling ten letters, not just the 5-5 ones discussed here.

I am much indebted to Steve Root for providing the computer expertise that made these calculations possible.