

Compass Points in English Surnames

Christopher K. Starr

University of the West Indies

West is by far more frequent as an English surname than any of the other compass points, followed in sequence by *North*, *East* and *South*. It has been hypothesized that this is due to the pattern of internal migration in England during the period when surnames were becoming established. An alternative is that this is a matter of fashion, the west of England being seen as exotic and adventurous at that time. These hypotheses are tested using a large data set that considers both the four names mentioned and a variety of compound names that incorporate the four compass points. The evidence does not favor either hypothesis. Rather, it is consistent with the null hypothesis that the adoption and retention of one or another compass-point surname is largely a random process.



Introduction

English surnames often consist of or incorporate the four compass points. Such names are mostly of two types, although with frequent conflation and confusion between them: a) a topographic name makes reference to a landmark or other local feature, while b) a habitation name is taken from a named locality. For convenience, we can refer to the two together as compass-point surnames.

It is a striking peculiarity of the four simple compass-points surnames (henceforth simple names) that they are not equally represented in the population. As seen in any British or North American phone directory or census, *West* is by far the most common, followed by *North*, while the *East* and *South* families are comparatively rare (Lindsey 1956). Examination of a variety of North American phone directories shows considerable consistency between cities.

In response to Lindsey's inquiry as to the cause of this pattern, Smith (1956) set forth a hypothesis based on two observations: a) simple names commonly arose as indicators of origin, and b) during the centuries in

which surnames became established in England, internal migration was predominantly from the west and north into London and surrounding counties. As an alternative, I suggest the hypothesis that during that period the west and north of Britain held a certain frontier mystique for the mass of English toward the southeast, so that the names *West* and *North* conveyed a certain cachet and were more likely to be adopted and retained. We can refer to these as the origin and fashion hypotheses, respectively.

Nifty as these hypotheses are, the relative numbers of simple names in the population provide no convincing support for either one, as the various instances of each name are not statistically independent. That is, while it is probable that a name that is now more prevalent had a greater number of independent origins, one cannot be reasonably certain of this in any single comparison.

Fortunately, simple names are not the only indicators of direction of origin in the case of internal migration. There are at least three distinct suffixes that at least sometimes serve this purpose. The two hypotheses both predict a significant common tendency in the ranks of the different compass points among the four sets of names together. (A “set” of compass-point surnames is defined here as all those sharing the same suffix or variations on it.)

However, such a tendency would be consistent with both the origin and the fashion hypothesis. In order to separate them, we have recourse to the many sets of compound compass-point names that usually do not reflect direction of origin. To give two examples: *Eastcott* could mean “dwelling in the eastern cottage” or could refer to a place by that name, while *Westbrook* could mean “from west of the brook,” “by the western brook” or could refer to a place by that name. There is no reason to expect that eastern cottages or western brooks should be more or less common than those in other directions. However, if some compass points were more fashionable than others in the period of surname establishment, a family dwelling in an eastern cottage or by a western brook might be more (or less) likely to have this reflected in its surname than one living elsewhere.

Accordingly, the origin hypothesis predicts a random pattern of ranks among non-origin names, while the fashion hypothesis predicts distinct general inequality among compass points.

Methods

My primary source of data was PhoneDisc USA (Digital Directory Assistance 1994), a directory of 81 million residential listings in the United States in the form of two CD-ROMs. (It is assumed here that the proportions of different English surnames among American telephone subscribers are an unbiased sample of English surnames worldwide.)

I examined two auxiliary sources of information on surname frequency in the U. S. Hanks (2003) provides frequencies from an estimated sample of one-third of the population, and the U.S. Census Bureau (1995) gives the abundance ranks of the 88,799 most common American surnames from the 1990 national census. The data-set in each is less extensive than that of PhoneDisc USA, but the expected strong correlation in relative frequencies gives confidence in the validity of data drawn from PhoneDisc USA.

Extracting these data and the subsequent statistical treatment were relatively easy. The truly demanding part of this study was in deciding which names to treat and which to leave aside, as there are at least three significant sources of error: names can be anglicized from similar Germanic forms, separate names may become conflated, and in some cases there is little certainty of a name's origin. I have attempted to minimize the chance of error by eliminating name sets dominated by one dubious name. My main sources in groping through this perilous landscape have been Hanks (2003), Hanks and Hodges (1988) and Reaney and Wilson (1997).

Any habitation name entirely or mainly based on one place is set aside as a source of bias. This refers in particular to English counties. For example, the existence of people surnamed Norfolk and Suffolk, but apparently not Eastfolk or Westfolk (or variations) requires no more explanation than the presence/absence of counties in England by those names.

Finally, I have omitted any set in which only one compass point is represented or in which the summed frequency is less than 100.

In computing ranks, I counted two compass points in a set as equally numerous if there was no more than a 10% difference between them.

Results and Discussion

As seen in table 1, the 81,000 simple names in PhoneDisc USA are close to the pattern noted by Lindsey (1956). *West* accounts for 66.5% of the total, *South* for a mere 3.9%. The prevalence of *East* is much greater than expected simply because Lindsey did not count such variants as *Eastes*, *Estess* and especially *Estes*.

At the same time, neither Smith's origin hypothesis nor the fashion hypothesis can derive much satisfaction from table 1 as a whole. Treating the different sets as if they were "observers" ranking the relative importance of the different compass points, we calculate Kendall's coefficient of concordance as $W=0.48$. This low figure indicates that the agreement among sets is quite weak.

Table 1. Numbers of subscribers in PhoneDisc USA with English compass-point surnames referring to direction of origin. Each suffix is given in its commoner form(s).

Suffix	East	North	South	West	Total
None	17,698	6302	3134	53,842	80,976
-ern	4127	815	1885	2668	9495
-man	5342	6	7	726	6081
-rich--ridge	1020	26,017	0	2	27,037
	28,187	26,017	5026	57,236	123,589

Table 2 shows the numbers of compound names in 24 sets that usually do not refer to direction of origin. As in table 1, there is tremendous variation within most sets in the relative prevalence of the different compass points today. The overall ranking of the four compass points is summarized in table 3. Here, too, there is a great deal of variation, but no strong general pattern is evident. This is seen in the mean ranks of the different compass points, which are remarkably similar to each other. Kendall's coefficient of concordance among the 24 sets is extremely low ($W=0.022$), indicating that the overall pattern is effectively random.

Compass Points in English Surnames 347

Table 2. Numbers of subscribers in PhoneDisc USA with English compass-point surnames that do not refer to direction of origin. Each suffix is given in its commoner form(s).

Suffix	East	North	South	West	Total
-brook	1430	1	15	6254	7700
-burn	535	3	0	0	538
-bury	2	177	220	288	697
-by	190	750	170	982	2092
-cliff	0	14	1470	0	1484
-cott~-cutt	301	2588	84	3547	6520
-ey	763	589	58	15	1425
-field	0	47	9	309	365
-gate	32	1	235	614	882
-hall	943	3	867	311	2124
-ham	674	528	170	20	1392
-hard	0	32	2970	3	3005
-lake	166	0	0	1008	1174
-land	345	145	10,845	213	11,548
-ley	5065	93	3	5805	10,966
-more~-moor	7	40	1	55	103
-over	0	45	0	1358	1403
-rop~-rup	5	3892	0	148	4045
-ton	3316	20,760	25,125	8947	58,148
-ward~-worth	2	8	2078	2	2090
-way	2	451	9	0	462
-well	0	35	598	16	649
-wick	107	130	1185	3	1425
-wood	1618	5849	261	602	8330
Total	15,566	36,118	46,383	30,500	128,567

348 Names 55.4 (December 2007)

Table 3. Rank distribution of compass points in 24 sets of English compass-point surnames that do not refer to direction of origin, derived from table 2. The figures show the number of sets clusters for which the compass point was the most abundant (1), second-most abundant (2), etc. Half-figures result from ties in some sets.

Rank	East	North	South	West
1	4.5	3.0	6.5	10.0
2	4.5	14.0	4.5	1.0
3	8.5	2.5	6.0	7.0
4	6.5	4.5	7.0	6.0
Mean	2.71	2.35	2.56	2.38

In conclusion, neither the origin nor the fashion hypothesis of the observed variation in prevalence of compass points is upheld. Much as we may regret it, the data are consistent with the null hypothesis that the adoption and retention of compass points in English surnames is a largely random process.

Acknowledgements

Rolf Bettner, Fred Cooper, Paul Garon, Maggie Kelly, John Krueger, Bob Moser, Pierre Sales, Justin Schmidt, Lynne Shivers, LaDonna Smith, Roy Snelling, John Solt, Bob Throckmorton and especially John Wentzel for bibliographic assistance. Carl Masthay, Anna Partington, Bob Throckmorton and (at all stages) Ed Lawson helped with onomastic/linguistic advice. Thanks to all.

Notes

1. Dedicated with respect and gratitude to Edwin D. Lawson, my teacher in the subject.

2. The raw data on which my analysis are based are available through December 2010 at <http://www.sta.uwi.edu/fsa/lifesciences/cstarr.htm> and at all times on request.

References

- Digital Directory Assistance. 1994. *PhoneDisc USA*. Marblehead, MA: Digital Directory Assistance.
- Hanks, Patrick. 2003. *Dictionary of American Family Names*. New York: Oxford Univ. P.
- _____, and Flavia Hodges. 1988. *A Dictionary of Surnames*. Oxford: Oxford Univ. P.
- Lindsey, D. 1956. "North and South." *Names* 4: 59.
- Reaney, P. H. and R. M. Wilson. 1997. *A Dictionary of English Surnames*, 3rd ed. Oxford: Oxford Univ. P.
- Smith, Elsdon. 1956. "West North Versus East South." *Names* 4: 166-67.
- U. S. Census Bureau. 1995. [Ranked list of surnames from 1990 national census.] <http://www.census.gov/genealogy/names/dist.all.last>