Long Term Trends in the Frequencies of Given Names

Douglas A. Galbi
Federal Communications Commission

The frequency distributions of personal given names offer important insights into the nature of the information economy. Here I present data on the popularity of the most frequent personal given names in the United Kingdom over the past millennium. The data show that the popularity of names, like the popularity of other symbols and artifacts associated with the information economy, can be usefully viewed as a power law. The frequency distribution of personal names, graphed as the logarithm of name popularity against the logarithm of name popularity rank, is similar to other popularity distributions where people and organizations are free to create and choose among many collections of symbols used in a similar way. Naming is seen to be representative of more general patterns of behavior in the information economy. Furthermore, the data suggest that historically distinctive changes in the information economy occurred in conjunction with the Industrial Revolution.

Names and the Information Economy

An important component of the information economy consists of the production and use of symbols. Names are an important type of symbol and choosing a "good" name involves an assessment of the social valuation of the name. The frequency distribution of names provides evidence of their social valuation and can offer insights into the historical development of the information economy. The frequency distribution of a name indicates the number of people who share the experience of being called by that name.

Since the early 19th century, the distribution of personal given names in the UK has evolved differently that it did over the previous eight centuries. Simple indicators of this change are the trends seen in the relative frequency of the most popular name, the three most popular names, and the ten most popular names. These show practically no

Names 50.4 (December 2002): 275-288
ISSN:0027-7738

differences from about 1300 to 1800. Since then all of these measures have changed dramatically, the result being a flattening in the name frequency distribution, viewed as a graph of the logarithm of name popularity against the logarithm of name popularity rank. This change in the evolution of the name frequency distribution early in the 19th century is suggestive of a more general change in the information economy about that time.

Popularity of the Most Frequent Names

Measuring name frequencies in actual samples requires attention to name definition and standardization. Given names can include multiple names and name variants as well as abbreviations, non-standard spellings, and likely mistakes in recording (e.g., William, Bill, Wm., Williamus). Unlike sampling variability, coding variability does not fall with sample size. Throughout the analysis here, names have been truncated to the shorter of either the first eight letters of the name or the letters preceding the first period, space, hyphen, or other non-alphabetic character (e.g. Wm. and Williamus would be truncated to Wm and William, respectively). These shortened names have then been standardized through a name coding process available on the internet (GINAP, by which, for example, Bill is standardized to William).[1] This procedure attempts to identify feasibly and consistently names with common communicative properties.[2] Experience with different name samples suggests that this procedure can reduce coding variability to less than half of one percent for the popularity of a single name and less than three percent for total popularity of the top ten names (Galbi 2001, Sec. I.B. and Appendix B).

Over the past two hundred years, the relative popularity of the most frequent names given in the UK has steadily declined. Table 1 shows popularity statistics for the most frequent names since 1800. The data in this table come from census records, birth records, and doctor registrations, collected from the sources documented in detail in Galbi (2001). In the UK from 1800 to 1994, the frequency of the most popular female name fell from 23.9% to 3.4% and that of the most popular male name fell from 21.5% to 4.2%. The popularity of the ten most frequent names for females fell from 82.0% to 23.8% and for male names from 51.5% to 28.4%.

Table 1. Frequencies of Most Popular UK Given Names, 1800-1994.

	Females				Males			
Birth	Top 1		Top 3	Top 10	Top 1		Top 3	Top 10
Year	Name	%	%	%	Name	%	%	%
1800	Mary	23.9	53.2	82.0	John	21.5	51.5	84.7
1810	Mary	22.2	50.7	79.4	John	19.0	47.0	81.4
1820	Mary	20.4	47.7	76.5	John	17.8	44.9	80.4
1830	Mary	19.6	45.4	75.8	John	16.4	42.3	78.2
1840	Mary	18.7	43.2	75.0	William	15.4	40.3	76.0
1850	Mary	18.0	41.0	72.1	William	15.2	38.7	73.8
1860	Mary	16.3	37.0	68.3	William	14.5	36.2	69.8
1870	Mary	13.3	31.5	61.1	William	13.1	31.7	63.5
1880	Mary	10.6	25.4	53.8	William	11.7	28.5	58.9
1900	Elizabet	7.2	16.2	38.5	William	9.0	22.9	50.9
1925	Mary	6.7	16.8	38.7	John	7.3	17.6	38.0
1944	Margaret	4.5	12.6	31.7	John	8.3	20.7	39.9
1954	Susan	6.1	13.2	32.5	David	6.3	17.4	37.8
1964	Susan	3.6	10.3	28.6	Paul	5.6	15.9	39.4
1974	Sarah	4.9	12.3	28.0	Mark	4.6	12.5	33.1
1984	Sarah	4.1	11.0	27.3	James	4.3	11.8	32.3
1994	Emily	3.4	8.6	23.8	James	4.2	11.0	28.4

Based on Galbi (2001), table 3 and underlying data. See appendix D for sources.

Before the 19th century, the frequency of the most popular given names in the UK was more stable. Tables 2 and 3 provide evidence on name popularity from late in the 11th century through the early 19th century. From 1300 to 1800, frequencies of 20%, 50%, and 80% are roughly typical for the most popular name, most popular three names, and most popular ten names for both males and females. The corresponding figures for the late 20th century are much lower—about 4%, 10%, and 25%, respectively. It is important to note that while top name popularities show no overall trend from 1300 to 1800, the particular names that made up the most popular names from one era to another did change. The causes of changes in individual names goes beyond the scope of this report, but for interested readers the names can be found in the appendix.

Table 2. Frequencies of the Most Popular Given Names in England to 1825.

	Females				Males			
	Top		Top 3	Top 10	Top		Top 3	Top 10
Year & Location	Name	%	%	%	Name	%	%	%
1080 Winchester					Robert	6.6	18.08	35
1120 Winchester					William	6.6	15.8	30
1180 Winchester					William	10.2	29.2	57
1200 Essex	Alice	11.3	27.4	56	William	12.4	30.5	61
1210 South	Matilda	16.2	39.9	70	William	14.4	32.7	65
1270 Rutland	Alice	19.4	51.0	84	William	15.2	35.6	76
1300 Lincoln	Alice	17.1	42.4	75	John	22.7	52.2	79
1260 London					John	17.6	39.7	69
1290 London					John	23.3	44.8	73
1510 London					John	24.4	49.4	74
1610 London					John	21.0	43.8	72
1825 London	Mary	19.2	43.9	73	William	16.3	39.2	80
1350 Manchester					John	22.7	47.5	92
1610 Manchester					John	18.6	37.6	77
1805 Manchester	Mary	25.8	48.2	84	John	21.7	48.7	81
1350 Yorkshire	Alice	22.4	50.4	86	John	33.5	66.8	93
1620 Yorkshire	Ann	24.0	54.7	88	John	16.2	47.1	86
1670 Yorkshire	Ann	21.5	59.2	79	William	18.7	47.4	78
1720 Yorkshire	Mary	25.7	57.4	87	John	25.5	57.8	86
1770 Yorkshire	Mary	22.8	45.9	84	John	25.6	55.7	86

1825 Yorkshire	Mary	20.1	45.8	81	John	18.8	44.2	79
1350 North/Cumbria					John	34.5	64.6	89
1530 North/Cumbria	Jane	16.0	44.8	84	John	23.1	46.1	74
1550 North/Cumbria	Margaret	15.6	45.1	86	John	21.7	44.4	75
1580 North/Cumbria	Margaret	16.8	44.9	84	John	18.0	39.4	71
1610 North/Cumbria	Elizabet	15.8	43.8	84	John	18.2	42.4	74
1640 North/Cumbria	Elizabet	16.6	46.0	87	John	19.7	46.7	75
1670 North/Cumbria	Elizabet	16.5	45.1	86	John	19.6	46.7	75
1700 North/Cumbria	Ann	16.4	47.1	86	John	21.1	49.5	77
1730 North/Cumbria	Ann	18.1	50.0	87	John	21.6	49.9	80
1760 North/Cumbria	Ann	18.8	52.1	89	John	23.2	51.4	81
1790 North/Cumbria	Mary	19.4	50.8	89	John	23.4	52.9	83
1825 North/Cumbria	Mary	20.3	46.7	88	John	21.8	49.9	85
1350 Hereford	Alice	21.9	47.2	84	John	34.8	58.9	89
1700 Hereford					John	20.3	49.9	78
1825 Hereford					John	18.9	51.1	90
1280 East Anglia					John	22.3	47.1	74
1400 East Anglia					John	36.1	63.3	90
1385 Soldiers					John	28.1	58.3	84
1550 Sailors					John	21.4	40.8	70
1560 Canterbury	Elizabet	13.6	33.6	74	John	20.3	46.9	75
1560 Glouster	Joan	18.7	45.5	88	John	21.4	52.5	80

For information on sources and sample sizes see Galbi (2001).

Table 3. Given Names in England, 1570-1700. Combined Frequency of the Three Most Popular Names (in %).

Birth Years	Females	Males
1570-1579	41.0	48.5
1580-1589	36.2	47.3
1590-1599	41.1	50.6
1600-1609	38.2	48.8
1610-1619	38.8	49.9
1620-1629	41.3	49.3
1630-1639	45.1	48.5
1640-1649	46.7	49.3
1650-1659	50.1	49.0
1660-1669	47.5	48.0
1670-1679	50.3	50.3
1680-1689	51.7	49.2
1690-1700	52.1	51.2

Source: Smith-Bannister (1997, 150).

Significant social, political, and religious changes in England prior to 1800 seem to have had little effect on the overall distribution of name frequencies. Within a few generations of the Norman Conquest of England, most given names were those brought by the invaders, although there is no evidence that the Norman clergy or court compelled the adoption of Norman names. By about 1250 pre-Conquest names had essentially died out.[3] Yet by the middle of the 13th century, the distribution of name frequencies was more like that of 1800 than it was of the late 20th century.

Describing Name Frequencies

The frequencies of the the most popular names follow a general order that can be recognized graphically. Earlier reports on given name frequencies, such as those of Eschel (2001), Tucker (2001; 2002), and Galbi (2001) have recognized that associated graphs have a characteristic shape when the logarithm of name popularity is plotted against the logarithm of name popularity rank, which is the same as a graph of name frequencies except that the left axis is labeled in more easily understood units.[4] The graph typically is close to a straight line. This type of empirical regularity is called a power law and it describes the relative frequency of popularity of names. Hence a power law, in this case, describes a relationship between the popularity of the most frequent name, the three most frequent names, and the ten most frequent names.

Figure 1. Popularity Distribution of Female Given Names in the UK.

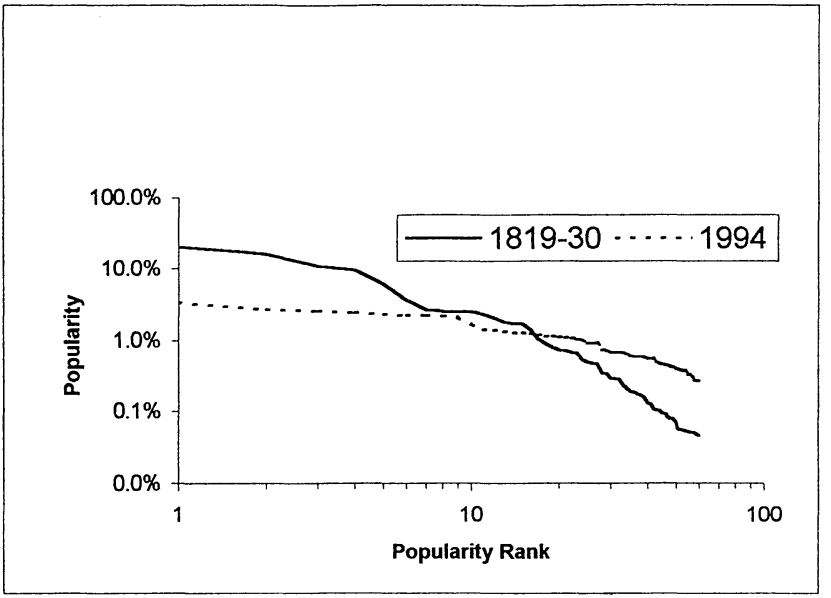
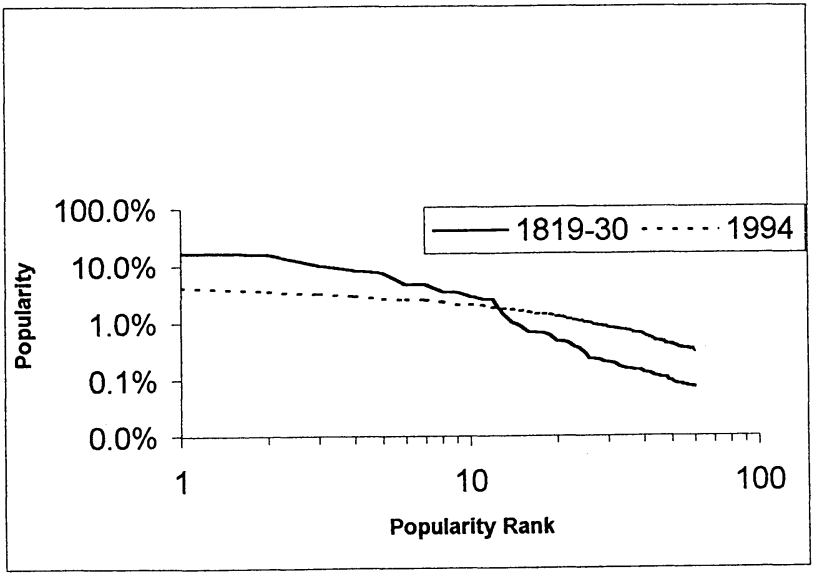


Figure 2. Popularity Distribution of Male Given Names in the UK.



Over the past two hundred years, the change in the popularity of the most frequent names has been associated with a flattening of the power law that best describes the name popularity distribution. Figures 1 and 2 show these graphs for names of females and males born in the UK in 1819-30 and in 1994. For both male and female names the slope of the line approximating the graph has become less negative, indicating that the relative name popularities have become more equal. This change can be interpreted as a reduction in the magnitude of information encoded in the name distribution and an increase in the extent of personalization in naming. (For a discussion of these points see Galbi [2001], Sec. II.B.)

Empirical regularities such as those seen in figures 1 and 2 are in fact prevalent in the information economy. Where persons and organizations are free to create and choose among many collections of symbols instantiated and used in a similar way, the relative popularity of the symbolic artifacts typically follows a power law. The circulation of magazines of similar type throughout the 20th century, the total box office receipts of movies, the popularity of musical groups as measured by gold records produced (Chung and Cox 1994), and the popularity of Internet web sites, measured in users or page views (Adamic and Huberman 2000) have all been shown to follow power laws. Insights into the evolution of such power laws from the study of name changes over time should contribute substantially to a more general understanding of personal preferences, media diversity, information industry structure, and other aspects of the information economy.

Conclusion

Although recent work on personal given names in England has emphasized name sharing in understanding the frequency distribution of given names (e. g., Smith-Bannister 1997), name sharing practices have little direct relationship to the frequency distribution of names. Naming children after parents, godparents, or ancestors is equally consistent with a high or low popularity of names. Similarly, having names freely chosen; that is, chosen in the absence of norms giving high value to the name of a person in a specific social position, could produce high or low popularity of certain names. The most that can be said for name sharing is that a norm of naming after parents creates additional inertia in name

popularity. Name popularity and its long-term evolution depend on factors other than name sharing. The evolution of the name frequency distribution over time is a complicated, dynamic system. Such systems can be highly sensitive to a particular factor at one time and totally unaffected by it at another time. Moreover, boundary conditions, such as a small share of naming done in violation of prevailing norms, can contribute to the overall state of the system.[5]

Analysis of long-term trends in personal given names in the UK suggests that significant changes in the information economy occurred in conjunction with the broad social and economic changes brought about by the Industrial Revolution, which began in Britain in the middle of the 18th century. The Industrial Revolution had a number of profound effects on English society. The population of England increased considerably and real economic income per capita increased by about a factor of four from 1300 to 1800, and by about a factor of 100 from 1800 to 2000.[6] It is not clear how the level of income might affect the frequency of names since populations of much different sizes show similar naming frequency distributions (Eshel 2001; Galbi 2001, table 4). The Industrial Revolution produced major changes in social networks and the social context of personal activity as well. The influence of these changes on the distribution of given names remains to be examined.

Whether information and communication technologies have created —or will create—a "new economy" is an important public policy issue. These technologies enable persons to interact in new ways that may bring about changes as significant as those associated with the Industrial Revolution. Consider, for example, the creation of knowledge about aggregate patterns of personal given names. Large compilations of name frequencies can be easily shared on the Internet. I have benefited from such sharing of information in writing this article, and I have made much more extensive data on name frequencies available through AGNAMES.[7] If other scholars use the Internet in similar ways, this subfield of onomastics could develop much more rapidly than it has in the past. The same might be true of many other areas of activity. Analyzing the popularity distribution of personal given names thus offers a particularly rich means for understanding changes in the information economy.

Appendix[8]

Most Frequent Male Names Given in London, 12th - 20th Centuries

c.1120[a]		c. 1260[b]		c. 1510[c]		c.1610[d]		c.1825[e]		c. 1994[f]	
Name	%	Name	%	Name	%	Name	%	Name	%	Name	%
Willelm	6.6	John	17.6	John	24.4	John	21.0	William	16.3	Jack	3.2
Robert	5.0	William	14.4	Thomas	13.3	William	11.4	John	13.5	James	3.1
Ricard	4.2	Robert	7.7	William	11.7	Thomas	11.4	George	9.4	Daniel	2.2
Radulf	3.6	Richard	7.0	Richard	7.3	Richard	5.2	James	8.6	Thomas	2.2
Roger	3.2	Thomas	5.3	Robert	5.6	Samuel	5.0	Thomas	8.6	Michael	1.6
Herbert	2.2	Walter	4.4	Ralph	3.3	Henry	4.8	Henry	7.6	Alexander	1.5
Hugo	1.8	Henry	4.1	Edward	3.0	Edward	4.5	Charles	5.8	Matthew	1.4
Johannes	1.3	Adam	3.1	George	2.1	James	3.5	Joseph	3.7	Luke	1.3
Anschetill	1.1	Roger	2.9	James	1.9	Joseph	2.6	Edward	3.5	Samuel	1.3
Drogo	1.1	Stephen	2.3	Edmund	1.6	Robert	2.4	Robert	3.1	George	1.3
N	912	N	814	N	427	N	463	N	48275	N	51097

[a]From the Winton Domesday book for the year 1148 (Barlow et al. 1976, 187). No details regarding standardization are provided.

[b]From the London Subsidy Rolls of 1292 (Ekwall 1951, 35). Variants include Walter (Water), and Henry (Hanry, Hary, Herri).

[c]Compiled from baptismal names in the London parishes of St. Peter's upon Cornhill, St. Denis Backchurch; Christ Church Newgate, Kensington, St. Antholin, and Budge Row (Stewart 1948, 110). Stewart notes: "I have tried to ignore mere variations of spelling, but to count separately the different forms of the same name, such as Henry and Harry, Augustine and Austin. . . . Spellings have been normalized to conform with those of the King James Bible, or with modern usage for non-Biblical names" (note, 109-110).

[d]Compiled from Stewart (1948, 112) using the same procedures noted for the list of 1510.

[e]Compiled from the names of persons born between 1819 and 1830 and recorded in the UK census of 1881. The complete census is available from the Genealogical Society of Utah (1997).

[f]From a list of all males born in Greater London in 1994 who were registered with the National Health Service Central Register (Merry 1995, table 21). The list above groups Jamie with James and Jake with Jack.

Most Frequent Female Given Names in Northern England, 14th-20th Centuries.

c.1350[a]		c. 1530[b]		c. 1600[c]		c.1730[d]		c.1825[e]		c. 1994[f]	
Name	%	Name	%	Name	%	Name	%	Name	%	Name	%
Alicia	22.4	Jane	16.0	Elizabet	16.6	Ann	18.1	Mary	20.3	Rebecca	3.7
Agnete	14.9	Elizabet	14.7	Ann	16.4	Mary	1.7	Jane	13.5	Lauren	3.6
Johanna	13.2	Margaret	14.2	Jane	13.1	Elizabet	15.3	Elizabet	12.9	Amy	2.8
Emma	7.7	Agnes	8.2	Margaret	12.8	Jane	11.2	Ann	11.6	Laura	2.6
Elena	5.4	Isabella	7.9	Mary	9.9	Margaret	10.6	Margaret	9.4	Jessica	2.6
Isabella	5.3	Alice	7.7	Isabella	5.9	Isabella	4.8	Isabella	6.2	Rachel	2.4
Margareta	5.1	Ann	7.3	Ellen	3.6	Hannah	3.0	Sarah	5.8	Sophie	2.4
Cecilia	4.5	Ellen	3.6	Alice	3.6	Ellen	2.9	Hannah	3.6	Hannah	2.4
Matilda	3.6	Catherin	2.8	Dorothy	2.9	Dorothy	2.6	Ellen	3.3	Sarah	2.3
Juliana	3.3	Mary	1.9	Frances	1.9	Sarah	2.1	Catherin	1.6	Emma	2.0
N	1794	N	852	N	2888	N	2038	N	24857	N	17719

[a]From the assessment roll of the 1379 poll tax for Howdenshire Hundrend, East Riding, Yorkshire (Gwynek n.d.). The names were listed in Latinized forms in the ablative case. Variants and diminutives were combined as follows: Alicia (Alisia), Agnete (Augnete, Annya), Emma (Emmota), Margareta (Mergareeta, Marg, Marg . . ., Mar', Magota), Cecilia (Cicilia, Sissota, Syssota).

[b]From parish marriage registers in Northumberland and Durham for 1538-68.

[c]From parish marriage registers in Northumberland and Durham for 1650-70.

[d]From parish marriage registers in Northumberland and Durham for 1740-60.

[e]From persons born between 1819 and 1830 in Cumberland, Durham, and Northumberland, recorded in the U.K. census of 1881.

[f]Includes all females in Cumbria, Durham, Cleveland, Tyne & Wear, and Northumberland in 1994 who registered with the National Health Service Central Register (Merry 1995, table 5).

Notes

The opinions and conclusions expressed here are those of the author and they do not necessarily reflect the views of the Federal Communications Commission, its Commissioners, or any staff other than the author.

1. Available at http://users.erols.com/dgalbi/names/ginap.htm. The principle for coding is to group together names that either sound the same, have the same public meaning, or changed only in the recording process (spelling errors, recording errors, etc.).

2. Among other things, name standardization helps to control for changes in names used as a person grows older, e.g., from Bobby to Bob to Robert. For this reason name standardization is particularly important in analyzing time trends when the data come from naming cohorts constructed by age. This is the case for data presented here on 19th century names.

3. There is no evidence that Norman clergy or royal officials compelled the English to adopt Norman names (Clark 1992, 552, 558-562).

4. This is true because $\log(a/b)=\log(a)+\log(b)$. The logarithm of name frequency differs from the logarithm of name popularity only by an additive factor. Name popularity rank and name frequency rank are of course identical.

5. Gabaix (1999) shows that, when the appearance rate for new cities is not too high, it has no effect on the slope of the power law describing city sizes. If the appearance rate for new cities rises above a certain threshold, then the slope depends on the appearance rate. Cities can be analogized to name types.

6. For population and income statistics for 1700 and earlier, see Mayhew (1995) table I, and Snooks (1995), table 3.5. For current population statistics, see UK National Statistics, Key Population and Vital Statistics, http://www.statistics.gov.uk/statbase/Product.asp? vlnk =539&More=N. The large changes in the structure of the economy over the past two hundred years make estimating changes in per capita income uncertain. The figure of 100 is an estimate based upon my understanding of the literature on economic history.

7. http://users.erols.com/dgalbi/names/agnames.htm.

8. These listings show the 10 most popular male names in and about London, c. 1120 to 1994 and the 10 most popular female names from north England (Yorkshire, Cumbria, and Northumberland) c. 1350 to 1994. The years given are approximate birth years, estimated relative to the date of compilation and the probable ages of the persons in the compilation. The data come from a variety of sources, which used different and often not explicitly described methods of standardizing and grouping the names. Readers interested in additional name lists should consult Smith-Bannister (1997), appendix C, which lists at decade intervals the 50 most popular male and female names in 40 English parishes from 1538-49 to 1690-1700. Unfortunately Smith-Bannister does not give the frequency of specific names nor the sample sizes. The weights given to individual parishes in each decade sample apparently change, but details are not given.

References

Adamic, Lada A., and Bernardo A. Huberman. 2000. "The Nature of Markets in the World Wide Web." *Quarterly Journal of Electronic Commerce* 1: 5-12. Also available online at http://www.parc.xerox.com/istl/groups/iea/abstracts/ECommerce/webmarkets.html.

Barlow, Frank, Martin Biddle, Olof von Feilitzen, and D. J. Keene. 1976. *Winchester in the Early Middle Ages: An Edition and Discussion of the Winton Domesday*. With contributions from T.J. Brown, H. M. Nixon, and Francis Wormald. Oxford: Clarendon Press.

Chung, Kee H., and Raymond A. K. Cox. 1994. "A Stochastic Model of Superstardom: An Application of the Yule Distribution." *The Review of Economics and Statistics* 76: 771-75.

Clark, Cecily. 1992. "Onomastics." *The Cambridge History of the English Language, 1066-1476*. Vol. II. Ed. Richard M. Hogg. Cambridge: Cambridge Univ. Press.

Ekwall, Eilert, ed. 1951. *Two Early London Subsidy Rolls*. Lund: C. W. K. Gleerup.

Eshel, Amram. 2001. "On the Frequency Distribution of First Names." *Names* 49: 55-60.

Gabaix, Xavier. 1999. "Zipf's Law for Cities: An Explanation." *Quarterly Journal of Economics* 114: 739-67.

Galbi, Douglas A. 2001. "A New Account of Personalization and Effective Communication." Available at http://www.galbithink.org or http://www.ssrn.com.

Genealogical Society of Utah, Federation of Family History Societies [GSU-FFHS]. 1881. *Census for England and Wales, the Channel Islands and the Isle of Man* [computer file, SN: 3643] (Colchester, Essex: The Data Archive [distributor], 29 July 1997).

Mayhew, N. J. 1995. "Population, Money Supply, and the Velocity of Circulation in England, 1300-1700." *Economic History Review* XLVIII: 238-57.

Merry, Emma. 1995. *First Names: The definitive guide to popular names in England and Wales 1944-1994 and in the regions 1994*. (With support from Kay Callaghan and Chris Cotton). London: HMSO.

Smith-Bannister, Scott. 1997. *Names and Naming Patterns in England, 1538-1700*. Oxford: Oxford Univ. Press.

Snooks, Graeme Donald. 1995. "The Dynamic Role of the Market in the Anglo-Norman Economy and Beyond, 1086-1300." *A Commercialising Economy: England 1086 to c.1300.* Eds. Richard H. Britnell and Bruce M. S. Campbell. Manchester and New York: Manchester Univ. Press.

Stewart, George R. 1948. *Men's Names in Plymouth and Massachusetts in the Seventeenth Century*. Berkeley and Los Angeles: Univ. of California Press.

Tucker, D. K. 2001. "Distribution of Forenames, Surnames, and Forename-Surname Pairs in the United States." *Names* 49: 69-96.

_____. 2002. "Distribution of Forenames, Surnames, and Forename-Surname Pairs in Canada." *Names* 50: 105-132.