Lucky Names: Demography, Surnames and Chance.

Nadav Shnerb, Yossi Maruvka and David Kessler

Department of Physics, Bar-Ilan University, Ramat-Gan Israel 52900

Demography is the scientific field that studies the size, structure and dynamics of populations. Humans and animals are not just a set of objects; each individual is someone's child, and their parents had parents of their own. More generally, every creature represents a leaf or branch on his family tree. This article attempts to explain a number of properties of family trees, for the purpose of understanding the degree of interrelatedness between people, the genetic diversity existing in the population and the distribution of various surnames.

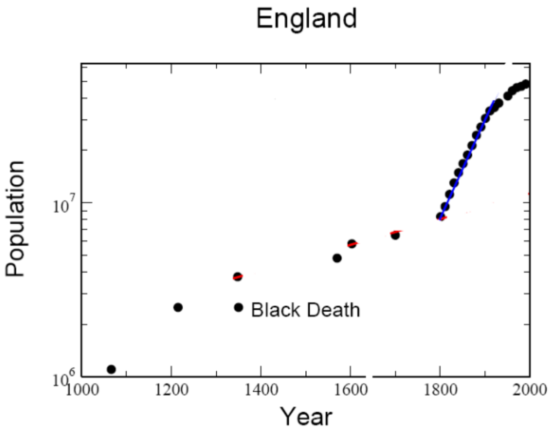
In the past, familial and tribal belonging carried great weight. A short glance in the Old Testament reveals that there are large sections dedicated to detailed lists describing the familial relations between people and groups. The question of familial belonging has shaped the course of history: conflicting claims as to Edward III's eligibility to inherit the French crown gave rise to the Hundred Years War; the Fatimid Caliphate, spanning the territory between Mauritania and Turkey, was built on the claim that Caliphs were direct descendants of Ali, holy to the Shiite sect of Islam; King Henry VIII withdrew from the Catholic Church and founded a new religion solely to enable himself to produce a male heir to inherit his throne. In the following sections some of these phenomena and their likelihood will be discussed. However, we will open with a fairly modern problem, namely overpopulation.

Thomas Malthus Foresees Disaster:

1066 was a momentous year in English history. A Norman adventurer named William landed on the shores of the Isle along with his confederates, and defeated King

Harold in the Battle of Hastings. They took control of the Island and crowned their leader William I, King of England, otherwise known as William the Conqueror (previously he had been known by the less flattering title "William the Bastard").

William and his army had a simple goal: loot. After overrunning England, his most important accomplishment was the execution of an extensive census of the population and their lands. The results of this were written up in a special volume, the Doomsday Book, which was kept in the king's palace in Winchester. The purpose of the census was to allow the efficient collection of taxes, but it provided future generations with detailed documentation of the size of England's population almost a thousand years ago. The other data points on the graph above are the results of subsequent censuses. Among other things, it is easy to spot the dramatic drop in the population after the bubonic plague pandemic which struck Europe in the middle of the 14th century.

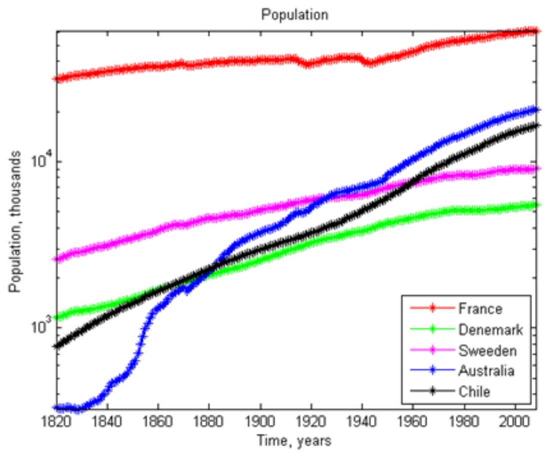


As can be seen on the graph, starting in 1800 censuses became more frequent; however, the most visible trend is the dramatic growth rate of the population since the beginning of the modern era. The data prior to 1800 show a growth rate of approximately 3.5 percent per generation, meaning that if in a given generation a thousand people lived in a certain town then 30 years later it had a population of 1,035. Meanwhile, after 1800 the growth rate jumped to almost 25% a generation, so that the population of our hypothetical town grew from 1,000 to 1,250 within the same 30 year period. This had obvious results. The plot of land inherited by each son upon his father's death became too small to sustain a family, since the lands were now divided among too many sons. Thousands of impoverished farmers flocked to the cities and began to work for starvation wages in various sweatshops. This gave rise to immense tension between the different social classes. The class struggle became the framework through which the people of the 19th century viewed their world.

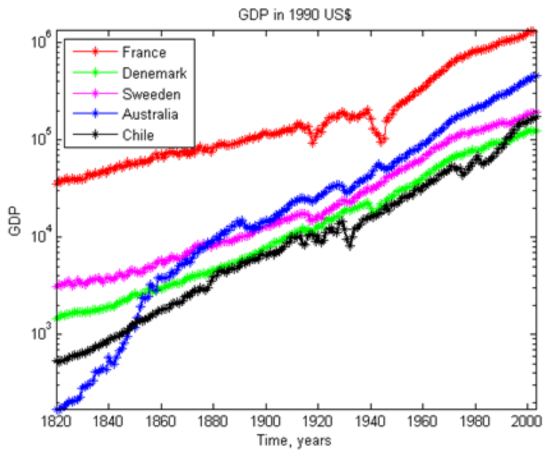
This population growth resulted from a drop in infant mortality rates alongside a rising life expectancy. These in turn resulted from a lower disease rate following improvements in public sanitation and a certain rise in the amount of available food. Despite this seemingly positive trend, there was at least one man who watched it unfolding with growing trepidation.

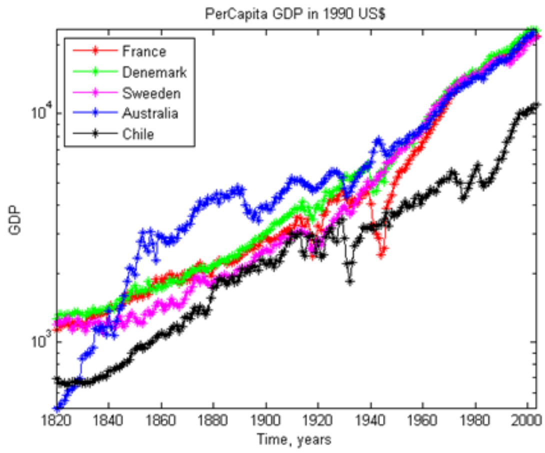
Thomas Malthus was a brilliant English demographer, economist, and Reverend who lived through the great demographic shift described above. He argued that there is no way to increase the production (of such things as food or housing) at a rate that would keep up with the growth rate of the population. Malthus reached the conclusion that there would come a time where there would not be enough food to sustain the population. Humanity would reach a point where some people would face starvation, logically leading to a series of wars and epidemics which would eventually cause a reduction in the population. Malthus did not see any other options. Either the birthrate must be halted (he recommended a variety of techniques, but in accordance with the Victorian morals of his time he emphasized voluntary abstinence), or the situation would reach crisis proportions. Because of this way of thinking, Malthus opposed the Poor Laws, Britain's first attempt at a welfare system. His claim was simple: if food or money is given to a poor person, preventing him from starving to death, he will bring four children into the world. Society will certainly be unable to take care of these offspring in the future, so it better to end the story in the earlier stage. "The Poor Laws", according to Malthus, "create the poor which they maintain."

It's difficult to say definitively that Malthus was wrong, but the fact remains that as of the present day his vision has not become reality. What has happened so far can be seen in the three pictures on the right which show the population (top), the GDP (middle) and the GDP per capita (bottom) for five countries from 1800 until 2000. It is clear that the population has grown at a good clip, but the GDP, or the amount of resources at the population's disposal, has grown **even faster**, meaning that the GDP per capita has risen as well. Additionally, following the dips caused by disasters and crises (such as the two drops that occurred in France – the red line in the graph – during World Wars I and II) equilibrium is quickly regained by excessive growth after the crisis, causing the overall upward trend to remain more or less stable.



The question arises whether Malthus was actually wrong, or whether at the end of the day the rapid growth rate will not be sustainable and a Malthusian catastrophe will occur. The prestigious journal *Science* presented this issue as one of the 25 most important problems facing modern science.[1] Some economists hold that the rate of economic growth in the modern era is a transient phenomenon, and that when all is said and done you can't fault Malthus's logic; food will run out if we continue to reproduce unchecked. Others are of the opinion that the prodigious growth of the GDP is a result of technology created by humankind. Therefore, the main resource at humanity's disposal is the creativity of individuals (albeit rare ones, but at some non-zero fraction of the population) which brings about new inventions and technological advances. If this is indeed the case then the central factor determining the rate of growth is the people themselves, and the more of them there are, the better (at least in the case of societies where people are able to engage in scientific and technological pursuits). The strategy suggested by Ted Baxter - having many children

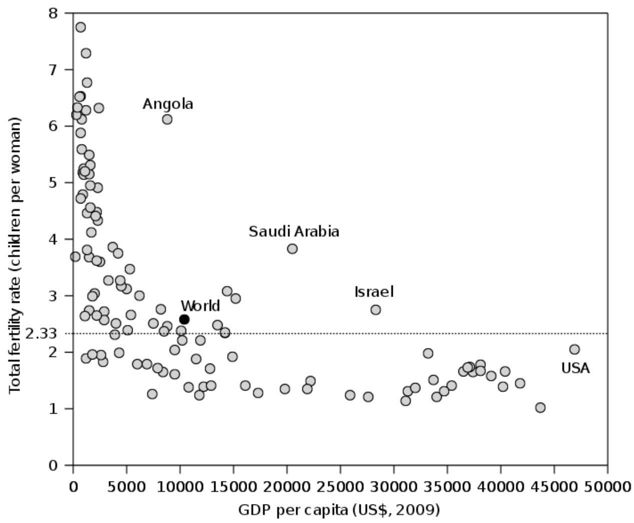




[1] Stokstad, E. 2005. "Will Malthus Continue to Be Wrong?" *Science Magazine* 309:102.

in the hopes that one of them would grow up to be a creative genius who could solve the overpopulation problem – may emerge as a realistic option.

It should be noted that over the last generation two opposing processes have been occurring. On the one hand, the eradication of contagious diseases has brought about a lower child mortality rate, even in places where large families and relative poverty are prevalent. On the other hand, education, technological sophistication and westernization are almost exclusively tied to a drop in the number of children per family. Today many European countries are experiencing negative population growth, which means that the number of deaths is higher than the number of births. This is even more apparent when you differentiate between new immigrants and the native born populations. The picture on the right provides an overview, showing the average number of births per woman on the y-axis and the GDP per capita on the x-axis for many different countries in 2009. In most of the wealthy countries the birth rate is lower than the critical rate of 2.33 children per woman (the dotted line). When the birth rate is below this point the population diminishes since some children will not reproduce from lack of desire or capability.[2]



Therefore, a Malthusian catastrophe does not appear to be imminent, so we can turn our attention to the next pressing question.

Who Wants to Be the Messiah?

"And there shall come forth a shoot out of the stock of Jesse, and a twig shall grow forth out of his roots…But with righteousness shall he judge the poor, and decide with equity for the meek of the land… And righteousness shall be the girdle of his

[2] For comparison, in Angola there are 6 births per woman, but a fifth of all babies die soon after birth and the life expectancy of the remaining individuals is 37 years.

loins, and faithfulness the girdle of his reins. And the wolf shall dwell with the lamb, and the leopard shall lie down with the kid; and the calf and the young lion and the fatling together; and a little child shall lead them … And it shall come to pass in that day, that the root of Jesse, that standeth for an ensign of the peoples, unto him shall the nations seek; and his resting-place shall be glorious." (Isaiah 11:1-10, JPS translation)

This maybe the best known of Isaiah's prophesies. Most of the biblical exegetes understood it to refer to the future redemption, and the man mentioned to be the future redeemer, the messianic king. If this is the case, then the verses indicate that the messiah must be "a shoot out of the stock of Jesse". The simple understanding of these words, as well as of the phrase "the root of Jesse" mentioned later, is that the messiah is supposed to be a descendent of Jesse. Who was Jesse? The obvious answer is that Isaiah is referring to King David's father Jesse the Beth lehemite.

This appears to be the only job requirement. All one needs to do to qualify to be the messiah is prove that he is a descendent of Jesse the Beth lehemite. For the purposes of further discussion, we will assume an additional constraint, namely that it is not enough to be just any descendant, but one must be the last in a line of sons: the son of a son of a son, and so on. What are one's chances of being a candidate for the position of messiah under these conditions?

Before examining the mathematical aspects of the question, two other facets should be introduced. First is the issue of surnames. In many traditional societies surnames are passed down from father to son, so that if Jesse had been the only man in his generation bearing the surname the Beth-lehemite, then under a traditional regime of name inheritance all of his male descendants would also have that very same surname. In modern-day Israel there are many people who in fact do bear this surname, but none, to the best of our knowledge, claim to have had it passed down to them through the generations from the time of King David. This reflects a common phenomenon whereby even in traditional societies people occasionally change their surname, either to a new name not previously found in their culture, or to an existing name. This may be done just because they liked the new name better than their old one, or they may perceive it to be advantageous. Thus Robert Zimmerman became Bob Dylan and Issur Danielovitch adopted the name Kirk Douglas. Therefore, after a number of generations there is no longer necessarily a connection between a surname and its bearers' ancestors.

Genetic inheritance, on the other hand, is not subject to the whims and wills of men. The nucleus of every cell in the human body contains genetic material in the form of chromosomes, half of which are inherited from the mother, and the other half of which are inherited from the father. All except for one piece, the Y chromosome, which all males carry and pass down to their male descendants. Therefore, this chromosome functions as a sort of immutable genetic "surname”. All of a man’s male descendants, generation after generation, will have the same Y chromosome as he did.

This statement is not strictly accurate. Although one cannot voluntarily change his Y chromosome, occasionally the body does not copy it completely accurately from the source, and the result is that a son inherits a Y chromosome which is not a hundred percent identical to his father’s. Each chromosome, or any other piece of DNA, can be described as a sequence of letters which makes up the genetic code. Sometimes one or more letters undergo a change, or in technical terms *mutate*, during the replication process. However, the scale is comparable to a few letters in a thick tome getting garbled, so that the difference in the Y chromosome from one generation to the next is not significant. Even the changes that accumulate over the course of many generations – and in our example approximately 3,000 years, or 120 generations at 25 years per generation, have passed – are generally small enough that for all practical purposes the Y chromosome remains practically the same. If this is the case, then what are the chances that one has the messiah’s Y chromosome? The fundamental piece of math needed for the answer was developed in a different context.

Where have all the British nobility gone?

On June 1st, 1858 papers by Charles Darwin and Alfred Russel Wallace were read at a meeting of the Linnean Society of London. These papers contained the core ideas of the theory that has been known ever since as Darwinian evolution, specifically, its basic thesis by which the development of life on this planet reflects a struggle for existence always won by the fittest, i.e. the strongest, the fastest, or the most clever of the participants. Since offspring tend to have qualities similar to those of their parents, a lion which happens to be born with stronger teeth or faster legs will not only be better at hunting zebras then his counterparts (leaving them with less food, and therefore less likely to produce young), but will also likely pass on to its descendants these superior qualities. Those qualities which improve the survival rate, cause more offspring to be produced, and are passed down genetically to the next generation are those which will persist, and the populations possessing them will be better able to compete successfully against weaker populations and destroy them.

Darwin's ideas caused enormous intellectual upheaval. As in any case where brilliant new ideas appear on the stage of history, people tried to apply this innovative way of thinking to numerous fields, both within its original context and outside of it. One of these directions was developed mainly by Darwin's cousin Francis Galton, an accomplished and multi-faceted scientist who worked on many fields ranging from statistics to photography. He promoted the idea of improving the human race called eugenics. The main idea of this theory is self-direction of human evolution. Specifically, the goal was to identify people with unfavorable qualities such as stupidity or criminal tendencies, and artificially reduce the number of offspring they produce by sterilization or preventing marriage. Correspondingly, there was a desire, at least in principle, to encourage desirable people – those more talented or stronger than average – to have more children. The scandalous implementation of baseless eugenic ideas by the Nazis during World War II drove the discipline out of fashion, but up until the 1930s very serious people invested their energy researching the field.

Galton, however, became worried. Studies showed that the surnames of the British social elite (judges, peers and the like) were diminishing. A significant number of prestigious surnames that had been a part of English society for centuries had disappeared by the middle of the 19th century. What was the meaning of this? Did the nobility have, for some reason, less reproductive capabilities? Many people were of the opinion that this was the case. John Stuart Mill, a prominent philosopher of the time, held that the nobility simply were not interested in large families, since then they would have to solve the problem of dividing the estate among their many children (a problem that does not arise in the case of the less fortunate). A physician named Thomas Jarrold suggested that the stress and tension that fill the lives of the upper class (who are afraid of losing their privileged status) impair their childbearing capabilities. An industrialist by the name of Doubleday claimed that the rich eat overly excessive amounts of meat, somehow reducing their fecundity.[3] But what was important to Francis Galton was the final result. If the stronger classes of society do have fewer children and the poorer, less fortunate have more, then "our population is chiefly maintained through the 'proletariat,' and thus a large element of degradation is inseparably connected with those other elements which tend to ameliorate the race."[4] A nightmarish scenario indeed.

There may yet be hope, however. Galton and his friend, the Reverend Henrey William Watson, asked themselves whether the disappearance of the "great names" may not be the result of insufficient reproductive abilities, but of simple happenstance. The basic idea is as follows: in a society in which people do not change their surnames, or at least do not change them to names that are already present within their community, a surname which disappears will never be able to reappear.

We can imagine a situation in which three males in a given populations share a surname. In the next generation the number of people with that surname is equal to the number of children of those three men. Thus, if by some unhappy circumstance two of those men never get married and the third man and his wife are unable to conceive then the surname will completely and permanently disappear. More generally, statistics tell us that the number of male sons a man has ranges typically from 0 to 4, and this number is distributed fairly arbitrarily. This guarantees that

[3] Robin Dunbar. 23.5.1985. "Is Every Theory Always Wrong?" *The New Scientist*, p. 44.
[4] Watson and Galton essay, http://galton.org/essays/1870-1879/galton-1874-jaigi-family-extinction.pdf

family size will vary randomly over the generations, causing there to be a certain probability that it reaches zero meaning that the family disappears.

In order to quantify the problem Galton and Watson discussed the following question: in a country with N_0 males with different surnames, where in every generation each of them produces a random number of sons (in order to keep the population at a fixed size, we will stipulate that the average number of offspring is one), assuming that every male inherits his father's surname and doesn't change it, what are the chances that a given surname survives over a period of time?

Their calculations showed an important result: the probability that a given name survives over a long enough time period is very close to zero. After enough time every citizen of the country will have a single surname, and all the other ones will be extinct. In other words, from the entire range of surnames in use today **only one** will survive over time.

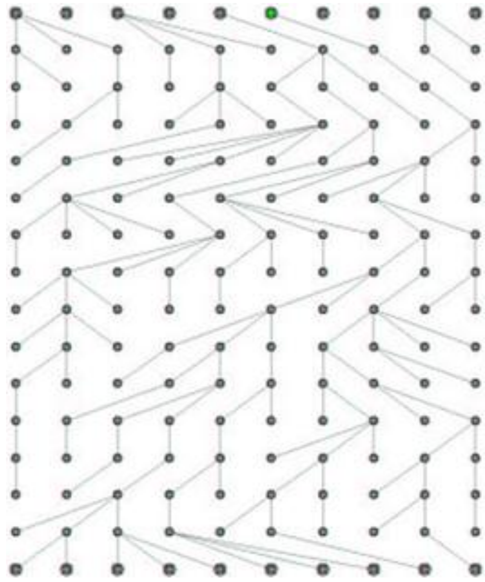
Galton and Watson's result is valid for the Y chromosome as well. Just as the theory predicts extinction for all surnames but one, it predicts that in a population of a stable size, given enough time only one man's Y chromosome will survive, and all the rest will disappear (this phenomena is known as "genetic drift"). The time required for this to happen is roughly equal to the size of the population (so that in a community of 100 people it will take 100 generations), but after S generations, there will typically remain only 1/S of the original surnames. These will appear much more frequently since there are fewer surnames for the same amount of people, so that each surname has more members. This means that if 150 generations have passed from King David's time to today, then there is a 1/150 chance that his line has survived, and if it has then it should be represented today by approximately 150 people.[5] These are the candidates for the position of messiah.

---

[5] More precisely, based on the Galton-Watson theory it can be shown that the chances of an individual man having S descendants after T generations drops exponentially [like $\exp(-S/T)$] for $S>T$.

This phenomenon raises questions about the large number of Cohanim (priests sect) among Jews. Priestly status is conferred through the male line, and recent genetic studies have showed that there is indeed a similarity in the Y chromosome of Cohanim from various places around the world. These have succeeded in identifying a segment of the chromosome as typical of the common ancestor of all

The Galton-Watson result can also be understood by looking at it qualitatively. Every person has one father, one paternal grandfather and so on. That means that every generation in his patrilineage consists of only one man. The chances of being patrilineally descended from Jesse of Bethlehem are the chances that your forefather in that generation, that one person out of all of humanity (or at least out of the entire Jewish people), was Jesse himself. People can be thought of as dots with lines between them connecting a father and his male descendants as in the picture on the right.



It represents a population whose size holds steady at ten individuals a generation. Every dot is a man, and every row is a generation. A father is connected to his sons by lines. Every son has a father (somebody has to sire him), but not every man has a son. The son's patrilineage is the line from him going upwards, and the chance of that line passing through any given dot is in this case one tenth. If the population under consideration numbers a million people the chances drop to one millionth. Therefore in an entire nation the chances are negligible.[6]

Returning to the original problem, has the line of Jesse, father of King David's reached an end? The answer is that there is almost no chance of finding a

---

Cohanim, Aaron the priest, brother of Moses (see Wikipedia on "y-chromosomal Aaron, http://en.wikipedia.org/wiki/Y-chromosomal_Aaron). Cohanim make up roughly 5% of the Jews living today, and attempts to estimate their numbers in 1800, prior to the rapid growth of the human population, have proposed that there were no less than 50,000 Cohanim at that time. The probability for a single man, living no more than 150-200 generations ago (assuming 25 years per generation), having that many descendants, is negligible in a stable population. This remains true even in a population that grows at a slow rate, such as was the case up until 1800.

[6] A similar problem appears for Shi'ite Muslims who believe that the leader in every generation should be an Imam descended through the male line from the fourth Caliph Ali Ibn Talib (the word Shia in Arabic means sect and the reference is to "the sect of Ali"). "Standard" Shi'ites, such as those in Iran, believe that Ali's line ended in the year 873, when the twelfth Imam, Muhammad the son of Hasan al-Aksari, disappeared off the face of this earth becoming the "Mahdi" (similar to a messiah), and ever since has been living in hiding until he reappears again as the future redeemer. Other Shi'ite sects accept this doctrine in principle, but dispute the identity of the disappearing Imam. Only one sub-sect of Shi'ites, the Nizaris, claims that their *current* leader (presently Aga Khan IV) is a direct patrilineal descendant of the Caliph Ali.

descendant of that pedigree if you look for a patrilineal heir. However it may be possible to look for an heir in other ways.

On Genetic Material and Spanish Surnames

The man in the picture on the right is the current king of Spain. He was baptized Juan Carlos Alfonso Víctor María de Borbón y Borbón-Dos Sicilias. The last part of his name reminds us that he is the last remnant of the house of Bourbon, the same European royal family which once ruled over half the continent and of whose politics Talleyrand is reported to have said “they had learned nothing and forgotten nothing”, not even the golden days when they ruled Sicily. The same cannot be said of Juan Carlos I, a man who received Spain as a dictator and shortly thereafter instituted democratic changes. Nevertheless, it seems that as far as surnames go, the Bourbons not only do not forget, but the names seem to pile up over the course of time. How did Juan Carlos come by his impressive name?

Among Spanish speaking peoples surname inheritance functions differently than in other European countries. Every person has two surnames, one from his father’s side and one from his mother’s. So if Juan Lopez Marcus marries Theresa Garcia Ramirez, their son would be named Mario Lopez Garcia. In all three cases only the first name is his given name, and the last two are surnames, first the father’s first surname and then the mother’s first surname. The accepted practice in Spain had always been to put the surname inherited from the father first and the one from the mother second, but in November 1999 Juan Carlos I signed a law allowing parents to choose the order in which their children’s names would appear in the interest of equal rights and prevention of gender-based discrimination.

The dynamics of Spanish surnames after the new law appear to be very similar to those of genes found in our DNA; every individual has two copies of every gene, and passes one of them (chosen at random) on to his children, so that every child has one gene received from his mother and one gene received from his father. In this fashion, if every couple has on average two children, a boy and a girl, then the names get dispersed among the population. A person has tens of thousands of genes, and each one of these functions like a Spanish surname – it goes on a journey wandering through his offspring – both male and female.

While a man's Y chromosome is only able to reach his descendants through the male line, other genes, which are not on this chromosome, such as the genes determining eye color, blood type and most other physical characteristics, can pass from a son to a granddaughter, from a daughter to a grandson and so on, anywhere in the family tree. While a male's Y chromosome is identical to that of a single man – his father's father's father going back one thousand or five thousand years, the rest of his DNA can contain one gene that originated in Julius Caesar and another that had its roots in Queen Esther. A man's genetic material is dispersed through all his various descendants and mixes with that of other people from his generation.

If we are interested in the entirety of human characteristics then there is no reason to focus specifically on the Y chromosome, or on any other gene for that matter. People feel just as close to their maternal cousins as to their paternal ones. In light of this, and in order to appreciate how closely we are all related to each other, it would be interesting to reverse the question and try to find all of our ancestors, both male and female, or to seek the source of all the genetic material found in a single person living today.

We will again consider the system qualitatively. Every person received his genetic material from two parents, a father and a mother, in the previous generation. Each one of them inherited their DNA from a pair of grandparents. This works out to mean that a given person's genetic material originated from four people that lived two generations ago, eight people from three generations ago, and so on. Accordingly, going back n generations, the number of people who contributed genetic material to someone of this generation is 2^n, making it a thousand ten generations ago, a million twenty generations ago, and a billion thirty generations ago. But thirty generations ago there didn't exist that many people on the entire planet. Does this mean that every single one of them is represented in our genome?

In fact, the number of our ancestors can be smaller than what was just calculated. For instance, assuming two cousins married each other, each one had two parents, but because two of those were siblings there were only three grandmothers and three grandfathers two generations before. It is easy to see that in a closed society such crossings will reduce the total number of ancestors.

The parallel mathematical problem was solved a number of years ago by two groups of scientists.[7] They have considered the following process: take a constant population of N_0 people, where every couple has two children on average in every generation, and the marital ties are not dependent on distance or on social status (that is, any male's chance of marry any given female is equal, and there are no distinction made between nations or cities in which people only marry each other). Assume for now that a person lives in a generation that for the sake of convenience we will label generation number 1. It turns out that after a certain number (S) of generations, one of the following two options will be true. **Either** a person has no descendants in the population (such as the case where he had no children or none of his offspring had children) **or** the entire population is descended from him, meaning that all N_0 people contain some piece of his genome.[8]

And what is this S, the minimal time to total mixing? It is a number that depends on the size of the population in the following way:

$$S=1.77\log_2(N_0)$$

For example, a population with a thousand members will reach total mixing after 18 generations. Every man whose genome has survived for 18 generations will have it represented in every member of the community. A population of a million people will get mixed in 35 generations, and a population of a billion after 53. If we treat the

---

[7] Manrubia, Susanna C., Bernard Derrida and Damián H. Zanette. 2003. Genealogy in the Era of Genomics. *American Scientist* 91:158.
Chang, Joseph T. 1999. Recent Common Ancestors of all Present-Day Individuals. *Advances in Applied Probability* 31:1002–1026.

[8] The term "piece of genome" may be misleading, but clarifying it might unnecessarily derail the discussion, so the explanation is presented in a footnote. The number of genes is finite - a person has about 30000 genes – so eventually one may be a descendent of a person that lived many generations ago without carrying any of his genes. Still, for long enough genome, all descendent carry at least a piece of sequence (not necessarily a gene, i.e., a part of the sequence that codes for proteins) of any ancestor.

human population on planet Earth as a constant population of a billion individuals (this is an approximation of the situation that held up until two hundred years ago), and assume a 25 year generation (generation length is supposed to equal the average age at which a woman gives birth), then we get 1,300 years (e.g. approximately the time between Muhammad and Churchill). If we take the Jewish people in the past as a closed social unit of a million people (again, not exactly the case, but a reasonable approximation), then we get a mixing time of 875 years.

As often happens with mathematical proofs, there are simplifications that don't hold up in reality, at least not perfectly. The theorem given above explicitly made the assumption that the chances for any male to procreate with any female in the population are all equal. It is clear that very few children are born to a father from New York and a mother from Kamchatka, or to a Nigerian and an Eskimo from Alaska. Nevertheless the approximate analysis performed on the general problem shows that the whole process of mixing is so rapid in any case that even serious geographic and social barriers cannot stop it, as long as the barriers can be (infrequently) breached.[9]

The conclusion, therefore, is that if what is looked at is not the line that describes a man's paternal lineage, but rather the tree that describes the set of all his ancestors, then the result branches out rapidly so that if a man who lived several generations ago is represented at all than we are all his descendants.

To summarize what has been said so far, the statement "X is descended from Y" has two possible meanings. The strict reading arises when X and Y are connected by a lineage made up of only one gender, son following son or daughter following daughter. In the case of males, the forefather and the descendants will have the same Y chromosome and, in a traditional society in which names are not changed, they will also share a surname. There is also a piece of the genome that passes from mother to daughter, and is not dependent on the father called mitochondrial DNA (mtDNA). This DNA is also found in males but they do not pass it on to their progeny.

[9] Douglas L. T. Rohde, Steve Olson & Joseph T. Chang. 30 September 2004. Modeling the recent common ancestry of all living humans. *Nature* 431:562-566.

So "strict" descent is the claim that a man (or woman) has the same Y chromosome (or mtDNA) as a specific person in the past.

The loose reading of "descent from Y" is the claim that there is some genetic path that passes through both males and females and connects someone alive today (X) to a person in the past (Y). Genetically, if a man is connected in the loose sense to an ancestor then his genome is identical in part (usually a very small part) to that of his forbearer, but (as long as they are only loosely and not strictly related) their Y chromosomes or mtDNA are not identical.

The overall results of the Galton-Watson and Derrida-Chang studies can be presented as follows: In a constant population it is very hard to be strictly descended from someone. The overwhelming majority of men who lived in the past have no living descendants through the male line, and the probability of this drops to zero the farther back in time you go. On the other hand it is very easy to be descended from someone in the loose sense. Almost all of the people once living who passed some of their genetic material on are represented in the genome of everybody alive today, since there is some line of descent connecting them to us.

Another way of looking at the same result is to differentiate between characteristics passed on to offspring of the same sex (from father to son and from mother to daughter) and those inherited by all of one's offspring. The former are destined to disappear from the population even if in the first generation they are passed on to several people. The latter can eventually "infect" the whole population[10].

What are the Kings of Judah Doing in Notre Dame Cathedral?

Almost everyone who comes to Paris passes by Notre Dame Cathedral on Île de la Cité in the heart of the city. This Gothic cathedral was built over several centuries and its history reflects, in a large part, that of France itself. During the French revolution, when Jacobin fanatics lead by Robespierre fought against all religious

[10] An interesting example is the line of secession to the British throne. The Act of Settlement of 1701 limited this list to the descendants (in the loose sense) of the Electress Sophia of Hanover (1630-1714). The number of living legitimate descendants has grown to about 5000 during 12-13 generations, and one may expect that within another 10 generations, more or less, all the British citizens will be potential heirs of the crown.

dogma, the cathedral was declared a place of worship of the "cult of reason" that they were attempting to establish. During that period not only were King Louis XVI, his wife and thousands of other people sent to the guillotine, but also the heads were chopped off of most of the statues of kings standing over the main entrance to the cathedral. The angry mob saw the kings of France, past and present, as the enemies and exploiters of the people, and the defacement of the statues was seen as a symbol of the fall of the Ancien Régime. The statues on view today have reconstructed heads; the original heads can be found in the Musée de Cluny in the Latin Quarter.

However, the Parisian mob was mistaken. They thought they were decapitating the Bourbons, but the twenty-eight statues on the front of the building actually depict the kings of Judah: David and Solomon, Rehoboam and Asa, Jehoshaphat, Jehoram and Uzziah and so on. Why are there statues of the kings of Judah guarding over a cathedral in the middle of Paris, and why do the faces of these same kings, painted by Michelangelo, look down on visitors to the Sistine Chapel in Rome? The reason is the same: it is not the position they held that brought about this honor, but the fact of their presence in the genealogy of Jesus Christ.

As we saw, the notion that the messiah is supposed to be a descendant of Jesse the Beth lehemite, or even of his son King David, became well entrenched based on the prophesies of Isaiah. The Christians also hold this position. It is not surprising, then, that the opening verses of the New Testament present the patrilineal descent of Joseph and Jesus from Jesse and David. In fact, the New Testament contains two such genealogical lists, one at the beginning of Matthew and the other in the Gospel of Luke. These two lists are not identical. The kings of Judah only appear in Matthew's version, while in Luke the path back to King David goes through people who did not make it to the throne, such as David's son Nathan, mentioned in 2 Samuel (5:13). Attempts to harmonize the lists, as well as explain various difficulties (such as there being sections with too few people covering certain long periods of history) have proved fertile ground for exegesis over the years.[11]

One of the major problems is caused by the doctrine of the virgin conception of Jesus by the Holy Spirit.

> These two Gospels declare that Jesus was begotten not by Joseph, but by the power of the Holy Spirit while Mary was still a virgin, in fulfillment of prophecy. Thus, in mainstream Christianity, Jesus is regarded as being literally the “only begotten son" of God, while Joseph is regarded as his adoptive father…
>
> The question then arises, why do both gospels seem to trace the genealogy of Jesus through Joseph, when they deny that he is his biological father? (Wikipedia)

One popular solution to this quandary is that Mary was also from the house of David. The relevance of Joseph's genealogy is then explained by the verse in Numbers "And every daughter, that possesseth an inheritance in any tribe of the children of Israel, shall be wife unto one of the family of the tribe of her father, that the children of Israel may possess every man the inheritance of his fathers" (36:8, JPS). This verse, as understood by the Christian exegetes, lays down an obligation to marry within your own tribe in order to keep the inherited land within the family. Accordingly, it is reasonable to assume that Joseph took a wife who was also among the descendants of King David, and therefore Jesus too is of the seed of Jesse the Beth-lehemite. The Catholic Encyclopedia summarizes this as follows:

> Tradition tells us that Mary too was a descendant of David. According to Numbers 36:6-12, an only daughter had to marry within her own family so as to secure the right of inheritance. After St. Justin (Adv. Tryph. 100) and St. Ignatius (Letter to the Ephesians 18), the Fathers generally agree in maintaining Mary's Davidic descent, whether they knew this from an oral tradition or inferred it from Scripture.

[11] For a summary, see *The Catholic Encyclopedia*, Genealogy of Christ newadvent.org/cathen/06410a.htm and Wikipedia: http://en.wikipedia.org/wiki/Genealogy_of_Jesus

However, in light of what we have seen above, this answer is mathematically problematic. Strict patrilineal descent is hard to achieve, and the chance of being thus descended from a specific historical figure is very small. On the other hand, if a looser definition of descent is permitted then it becomes much more likely. Specifically, if Mary was descended from King David in a lineage spanning, according to Luke's genealogy, 42 generations, then it is reasonable to assume that almost every Jew in that generation (and maybe also people from other nations, since in the days of the Second Temple there was much intermarriage between Jews and the neighboring peoples) is a descendent of King David according to this criterion, and then all of the detailed genealogies lose their value. From this perspective St. Augustine's explanation seems much more reasonable. He proposed that since Joseph was married to Mary, Jesus is considered his son for matters of pedigree.

Mutations, Surnames and the Theory of Evolution (Again)

We have shown that in a population with a fixed size, over time one surname will take over. Of course, there are no actual societies in which this has happened, since when this is the case surnames become worthless, but according to Galton-Watson on the way to this takeover there will be a stage in which a small number of surnames will appear with a high frequency in the population. This has caused social problems in a number of cases. In Korea 45% of the population has one of the three surnames Li, Kim or Park, while in China more than 90 million people bear the surname Wang, making finding phone numbers impossible, and causing the police to arrest the wrong people when chasing after criminals. The BBC reported[12] (12.6.2007) that the Chinese Government is trying to convince people to give their children original surnames, and therefore is proposing the option of giving children a surname formed by combining the father's and mother's surnames, or as one newspaper, The China Daily, reported it: "If a father's family name is Zhou, and the mother, Zhu, the baby could have four options for the surname: Zhou, Zhu, Zhouzhu or Zhuzhou".

Uhhhmmm….

[12] http://news.bbc.co.uk/2/hi/asia-pacific/6745259.stm

Well, we have to admit that we have no idea how to pronounce these names, but the Chinese attempt brings us back to the concept we defined earlier, both in the context of surnames and in the context of genetics: mutation. When speaking of characteristics passed from father to son, mutation is the case when a son does not inherit his father's surname, at least not exactly. In the case of the Y chromosome this is caused by mistakes in the genetic copying process. In the case of surnames there is generally a conscious human choice behind the change of surname, though occasionally it is brought about by a mistake in spelling when the bearer immigrates to a foreign country.

The Galton-Watson result cited above does not take mutations into account. It assumes that every son receives his father's surname and that the disappearance of surnames is caused by a randomly varying number of offspring in each generation that eventually reaches zero. Now we understand that there is also an ever-present loss factor: some people change their surname, which reduces the frequency of a given surname.

The other aspect of the influence of mutations is the constant formation of small families. In an imaginary scenario where only one person is allowed to change his surname every time, mutation means the constant creation of families numbering only a single person each. Therefore, if the Galton-Watson theory predicts that in a stable population a single surname will take over and the rest will disappear, when there are mutations this will never happen. New families are formed all the time, some of which disappear, and some of which develop into larger families.

When this situation is treated mathematically, it turns out that a closed society in which N_0 people pass on their surname to their sons, but once in a while someone changes his surname, an equilibrium will be reached in which there is a fixed distribution of family sizes. This means that there will be many surnames and there is a function n_m that tells us what the chances that a surname chosen at random (taken with equal probability from a list of all the current surnames, without taking each one's frequency among the population into consideration) will be the name of m people.[13]

The solution for the function n_m was presented by the mathematician Warren Ewens in 1972.[14] According to Ewens:

(1) $$n_m \approx \frac{e^{-\mu m}}{m}$$

where the parameter μ is tied to the mutation rate (the chance of any individual changing his surname). Qualitatively, the more widespread the tendency to change your surname is, the less we expect to see "mega-families". In Korea the most common surname, Kim, belongs to 22% of the population, while in China the most common surname covers 10%. In the United States, in contrast, there are only 2.5 million Smiths and 1.8 million Johnsons – less than one percent of the population. If we assume that the dynamics of surnames work as described above, then it is clear that the chances that a Korean changes his name are a lot smaller than those of an American to do the same.

But is this really the case? We don't actually know the answer to that. For all we know, it might be that in Korea people change their name to Kim because it allows them to push to the head of the line while waiting for the train. Maybe the name Wang is so common in China because people of that name pass on a genetic tendency to reproduce prodigiously, or maybe in previous generations these families were more conscientious about hygiene, and therefore their children were less affected by diseases. Many reasons can be invented for why any given surname succeeded better than others. Most people tend to think this isn't the case; that is, that the Smiths and Johnsons are in first place not because of mysterious procreative powers, but simply because of historical accident. This holds true as long as the topic of debate is surnames, but when it comes to other subjects, opinions to the contrary abound.

[13] Note the difference between the case where a surname is chosen at random and the case where a person is chosen at random and asked what his surname is. In the latter case the result will be slanted in favor of the more common surnames.

[14] Ewens actually solved a much more general problem, and the result quoted here is a special case of that solution.

The basic question, which comes up in many contexts, concerns the distinction between luck and fitness. It is often tacitly assumed that bigger equals stronger (or better, or smarter or more clever or more ruthless, depending on whether it is a widespread biological species, a long serving politician, a famous composer, or a wealthy businessman that is being discussed). The other option, which is that bigger equals luckier, is raised with much less frequency.[15]

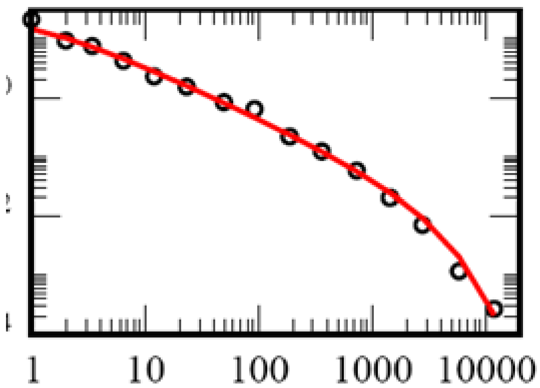
The most famous of fitness doctrines is of course Darwin's theory of evolution. The term *fitness* is central to Darwin's worldview. If there are two lions which are completely identical except that one of them runs faster than the other, then the faster one has a better chance of catching prey. Assuming that the amount of prey is limited and only enough for one lion, then presumably the slower lion will die. In any case, the faster one will be healthier and stronger and less hungry, so he will beat the slower, hungry lion in the competition for a mate, and will therefore, on average, produce more offspring. Since traits such as speed are hereditary (speed is related, for example, to the leg muscles), then chances are that the next generation will possess them. Over time, claimed Darwin, the slow lions will disappear, leaving only the fast ones. In the meantime, a new **mutation** can occur; a lion could be born with a heart defect, or a different color coat which provides better camouflage on the savannah. Harmful mutations (such as defective hearts) will disappear and useful mutations (such as camouflage) provide their bearers with better survivability, and will therefore persist and spread among the population. Darwin called this mechanism which allows evolution to take place **survival of the fittest**. In every population offspring are produced with a variety of qualities and the good ones remain while the bad ones disappear, causing the variety of characteristics to "flow" through the generations.

Comparing a property like "runs fast" to a property such as "surname" gives rise to both similarities and differences. Genetic mutations (such as the one that causes the lion to run fast), similar to surnames, can be assumed to be random. At the level of the genome, mutations occur because of chemical deficiencies in the process of

[15] Here is the place to mention Nissim Talb's books *The Illusions of Randomness* and *The Black Swan* which are dedicated to this question.

replicating the DNA, and people change their names for so many reasons that the final result also appears random.

The big difference between the two is with what happens after the mutation occurs. A reasonable assumption is that in most societies, in most time periods and for most surnames, the name itself is more or less a **neutral** property. A person's contribution to the dispersal of the surname (i.e. the number of children who will bear the name) does not really depend on whether his name is Kessler, Smith, Jones, or Cohen. On the other hand, according to Darwin a fast lion and its offspring (which can be thought of as having the surname "fast") will bear more young on average causing their frequency within the lion population to rise very rapidly. Darwin explicitly stated the assumption that "more common" equals "stronger" in his book, saying that "when we look at the plants and bushes along the riverbank, we're tempted to attribute their number and type to what we call luck, but what a mistake that is!"



Is this a mistake? Approximately a decade ago, ecologist Steve Hubble[16] took the scientific world by storm when he presented the graph below (this is not the original picture, but resulted from our processing the same results). The circles represent the probability of finding m individuals with an arbitrarily chosen surname. The graph shows that there are many small surnames (the circles on the left – there is a high probability that the number of people bearing that arbitrarily chosen surname is small) and a few large surnames (the circles on the right – large families, small probability). The red line that connects the circles is the theoretical prediction according to equation (1) above. It is clear that the match between the theoretical prediction and the actual data is very good.

[16] Hubbell, S. P. 2001. *The unifed neutral theory of biodiversity and biogeography*. Monographs in population biology 32. Princeton University Press, Princeton, N.J.
Volkov, I., J. R. Banavar, S. P. Hubbell, and A. Maritan. 2003. Neutral theory and relative species abundance in ecology. Nature 424:1035.

So what does this have to do with Darwin? The answer is that we lied to you. The circles on the graph weren't taken from surname statistics, but rather from a completely different field: frequency statistics of trees in a tropical forest. There is an island in the middle of the Panama Canal, Barro-Colorado Island, which has a section of jungle containing approximately 300,000 trees that are carefully monitored by the Smithsonian Institution. Among other things, each tree has been catalogued according to its species, and what can be seen in the graph is the probability that a species chosen at random is represented by m trees in that jungle.

Hubble claimed that if this is the case then there is no need for Darwin's concept of fitness to explain why a certain species of tree is common and another rare. From his point of view it can and should be assumed that all of the tree species have the same fitness; every species reproduces over the course of generations in a random Galton-Watson process and undergoes neutral mutations which create new species but which do not change their relative fitness. Biological species behave exactly like surnames, and the result is identical statistics. This provocative claim caused a deep shock in the ecological community and the vocal debate on the subject still fills the pages of the most prestigious scientific publications even as these lines are being written.

So surname statistics explain the relative frequency of trees. But where are the data on surnames themselves? Why have we not presented a graph, like the one above, that shows the frequency of surnames in a given society and their correspondence with (1)? The answer again is that we lied. In reality, surnames do not obey the "correct" statistics as produced using the Galton-Watson process. In order to understand the reason for this we need to revisit the problem that opened this article: the growth rate of the human population.

Surnames in a growing Population

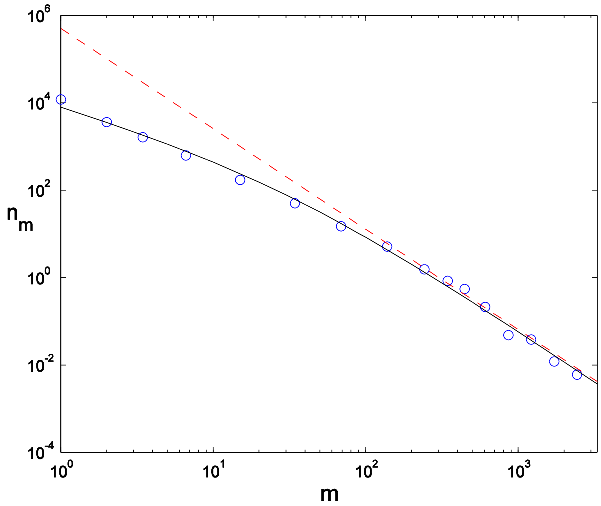
The data presented at the beginning of this article for the population growth rate in England sheds light on the reason the models presented above (those belonging to Galton-Watson and to Ewens) cannot explain the statistics observed in the frequency of surnames. The underlying assumption of all of those calculations was that they were dealing with a constant population of N_0 individuals. This assumption, while reasonable when dealing with trees in a rain forest, is simply incorrect when dealing with human societies. As we have seen, even in the society of pre-industrial England the growth rate was not negligible; in England the population grew from one million to ten million people over 700 years. By the standards of today this might not look significant, but compared to other biological species this is a dramatic increase.

How does this influence the frequency distribution? It is easy to understand if we return to the Galton-Watson model. They assumed that every man has on average one child who carries his name. This would be the case for example if the surname was passed through the male line, and every man in the population had equal chances of having two boys or two girls (and zero boys). Therefore, a surname held by a single person would in the next generation either be the surname of two people (with a 50% chance) or disappear (also with a 50% chance). Similarly, it is easy to be convinced that the chances that a family of a 100 would become in the next generation a family of 110 are the same as the chances that it would become a family of 90, or in other words that families jump randomly right or left on the scale of size. Since any family that, as a result of one of its random jumps, reached zero would disappear forever, it is clear that over time this will happen to all of the surnames, which is the basis for the result that says that all surnames eventually disappear. This is equivalent to the claim that a drunkard walking randomly either left or right and on whose left, at a given distance from him, is a chasm, will eventually fall into it, independent of his initial distance from the cliff's edge.

When the population grows, things work differently. If every man leaves three boys or three girls with equal probability then a family of one will find itself after one generation with an average size of 1.5, or alternatively, the chances of a family growing from 100 to 110 will be bigger than those of it shrinking to 90. This is a sort of middle ground between the drunkard walking randomly right or left and a sober man walking only right in order to maximally increase the distance between him and the chasm. It is still clear that a little bad luck might destroy small families, but large families are less vulnerable in this case and will presumably not die off.

We have recently developed a new mathematical tool[17] that solves this problem, building on earlier work by Susanna Manrubia and Damien Zanette.[18] We expand the original Galton-Watson model in two ways. First, there are mutations. Every man passes his surname on to a random number of children, but any one of those children can change it to something new. In addition, the size of the population is not stable. The average number of children a man has is larger (or smaller) than one, so that the population grows (or diminishes) (the case of a fixed population is of course a special case of our theory).

It turns out that in the end there are two parameters, the growth rate of the population and the probability that an individual will change his surname. Using both of these you can get a mathematical expression of the number of families, n, of size m.



How well does this work? This can be seen in the graph on the right, which shows the distribution of surnames as were documented in the 1790 census in the United States. We took the results of such an old census for two reasons. The first is that at that time the population of the US was mostly of English origin, so that the large waves of immigration had not yet brought with them the many foreign surnames currently prevalent among Americans. Second, we wanted a result from a population with a more or less steady growth rate, so that it could be treated as a constant, and we wouldn't need to take into account the demographic revolution of the modern era. It is clear that the model (solid black line) matches the empirical

[17] Maruvka YE, Kessler DA, Shnerb NM. 2011. The Birth-Death-Mutation Process: A New Paradigm for Fat Tailed Distributions. *PLoS ONE* 6(11): e26480. doi:10.1371/journal.pone.0026480

Maruvka YE, Shnerb NM. Kessler DA. 2010. Universal features of surname distribution in a subsample of a growing population. *Journal of Theoretical Biology*, 262: 245-256.

[18] Manrubia SC, Zanette DH. 2002. At the boundary between biological and cultural evolution: The origin of surname distributions. *Journal of Theoretical Biology* 216: 461–477.

results (blue circles) nicely. Extending the theory to include two growth rates we succeeded in achieving a nice match for the US census rates of 2000 as well.

These new insights allow us to say something about the results presented earlier regarding descent from the house of David in a growing population. It turns out that the idea that a trait that is inherited both by sons and by daughters, quickly takes over the population, is even truer for a growing population. This means that the claims made above are valid. In the case of a male line of descent there is a difference between the two cases but it is very small at the inconsequential rates of growth that we reconstructed for the human population.

In conclusion, the great interest in genealogy, which is sentimental in origin, brings us to an important scientific question: the statistical properties of a neutral process (with no fitness) of birth, death and mutation. This process has many interesting characteristics, some of which were reviewed here, while others, mostly those unrelated to the dispersal of surnames during immigration, have received much attention in other forums recently.[19] The most interesting question from a semi-philosophical standpoint is that of the relationship between deterministic factors (such as quality or fitness) and stochastic ones (luck) over the course of the history of the human race, other biological species and evolution in general. Is there a correlation between the number of offspring a man or woman has over the generations? Is what will determine the survivability of our offspring the genes that they inherit from us or blind chance? We do not yet have unequivocal answers to these questions, but with a little luck – and maybe some fitness - they will be reached in the course of future research.

[19] Korolev, K. S., Avlund, M., Hallatschek, O. and Nelson, D. R. 2010. Genetic demixing and evolution in linear stepping stone models. *Rev. Mod. Phys*. **82**:1691.