Mathematics knows things about physics that we don't (as yet)

Robert Shour

Abstract. Mathematical concepts compress solutions to problems involving natural and other phenomena. Hundreds of generations of human society have tested those solutions for their accuracy and utility and have continuously refined them. Those concepts likely often reflect fundamental aspects of the universe including phenomena not yet modeled, perhaps not yet identified. Mathematics therefore has predictive power. The accumulation of solved mathematical problems that we possess represents distilled intelligence that enormously exceeds that of any individual human being.

1 Introduction

In his 1959 lecture, *The Unreasonable Effectiveness of Mathematics in the Natural Sciences*, theoretical physicist and mathematician Eugene Wigner observed that "mathematical concepts turn up in entirely unexpected connections" particularly in physics and "there is no rational explanation for it". Computer scientist R. W. Hamming's 1980 update noted that there is "no rational explanation (as yet)". In 2005, Sundar Sarukkai observed that language itself is unreasonably effective and that mathematics, a collection of sub-languages, is "more suited to certain types of description, typically quantitative" than languages such as English. Why mathematics is so effective remains a live issue.

Mathematical reasoning itself (this article proposes) can explain, at least partly, the effectiveness of mathematics. First, in all languages, including mathematics, societies continuously improve the utility, efficiency and dissemination of concepts. To be effective, those concepts must be consistent with fundamental features of our physical environment. If our local physical environment is universally representative, then our local mathematical principles apply generally. Second, brains networked over time have more problem-solving capacity than individual brains. Third, the capacity of a network to solve problems is proportional to its degrees of freedom. Quantifying the degrees of freedom of a network enables calculation of how much more the problem-solving capacity of networked brains is than that of an average individual brain. If such an explanation exists, it may increase our appreciation of the efficiency of mathematics in the natural sciences.

2 Society creates and improves methods and concepts

On birth a child does not know physics, mathematics, how to read or speak, or what parents are. The child learns a culture and a language that existed long before it was born. Some 150,000 years ago or thereabouts our ancestors developed the idea of encoding information with vocalized sounds into a language. For over 100,000 years they collectively solved language problems that included identifying and choosing a set of compatible sounds to use for encoding, what should be encoded, and which sounds should be used for any particular encoding. Around 3,500 BCE our ancestors collectively began to develop writing. During the time that separates the beginning of speech and the beginning of writing, millions of people collectively, using their energy, added to their posterity's common store of solved problems. For a set of collectively solved problems, such as those in mathematics, "the inner logic of its development reminds one much more of the work of a *single* intellect, developing its thought systematically and consistently using the variety of human individualities only as a means" (I. R. Shafarevitch, in Davis, 1995, p. 56).

The phonemes available in a language together with perceptions, actions, concepts, etc. are data to which inference is applied to solve problems of word creation. Denote data, words, and when joined with inference based on them, if any, 'inference sets'. Inference sets are hierarchical in language, for example in categories (dog, mammal, animal), grammatical rules, and sets of words (phrases, sentences, books) and in mathematics (postulate, theorem, calculus or geometry). Inference sets are pervasive — smells, textures, sounds, recipes, morals, laws, metrics, apps, events, metaphors, stories, histories, athletic moves, theories, technologies, games — and hierarchical. Construction, engineering and manufacturing make inference sets physically manifest. Inference sets are collectively appraised by society for their utility and ease of use just as objects and services for sale in the marketplace.

Apply economic reasoning to a society's deployment of its collective energy resources — its problem solving resources — to improve the efficiency of inference sets: "...a potatoe-field should pay as well as a clover-field, and a clover-field as a turnip-field" (p. liv, the economist Jevons, 1879). "The product of the 'final unit' of labor is the same as that of every unit, separately considered" (p. viii, the economist John B. Clark, 1899). The invisible hand applies to the economics of inference sets; problems are like turnip-fields. By analogy, just as in an economy there are "dispersed bits of incomplete and frequently contradictory knowledge which all separate individuals possess" (p. 77, Hayek) leading to a market solution that "might have been arrived at by one *single* mind possessing all the information" (p. 86, Hayek), a society collectively and optimally allocates its problem solving energy resources. If, as the economist Friedrich Hayek observed, disconnected individuals can collectively find an optimal way to spend money on goods and services, they can also find an optimal way to spend energy on problem solving. The common currency is the estimated energy cost. Adoption by a society of a proposed solution to a problem depends on society's collective estimate of its learnability, utility and simplicity. Infer that the average rate of society-wide collective problem solving that improves inference sets should be same for all varieties of problems.

The history of mathematics, sciences and languages shows that inference sets improve over time. Analogize the increasing efficiency of language, for example, to the evolution of data compression software, which steadily improves data compression. A language increases its compression through naming (categorization and generalization), contraction (can't), clipping (bike for bicycle, bus for omnibus), acronyms (GI, UN), allusion (his, computer, WWII), pattern, metaphor, word order, word endings, gesture, musicality, and intonation ('yea, sure'). Mathematical notation compresses and encodes already compressed concepts and methods, as in the integral sign. The limit concept's history stretches from Eudoxus of Cnidus 2,400 years ago to modern times. A single person cannot duplicate that conceptual effort, but can obtain in a life span the advantage of our ancestors' and other people's energy invested in solved problems.

Software engineers collectively steadily make data compression software more efficient. Societies collectively steadily make inference sets more efficient.

3 Networked brains network neurons

Societies have greater resources to create and improve inference sets than individuals. Social insects and animals have 'swarm intelligence', an observation with a long lineage. Pappus around the year 300 attributed intelligence to bee hives making hexagonal cells (Vol. 2, Ch. XIX, p. 390, Heath). "Individually, no ant knows what the colony is supposed to be doing, but together they act like they have a mind" (p. 250, Strogatz 2003).

An ant has about one million neurons; a human brain has about 100 billion. Ten thousand ants in a colony have 10 billion neurons. That may explain why "together they act like they have a mind"; maybe networked neurons don't care where they are located when they exercise their degrees of freedom. If the same mathematical laws govern efficient allocation of energy within networks then for the same number of individuals a human society has as much more problem solving capacity than its average human as an ant society has than its average ant. A society of 100 million humans — 10,000 networked societies of 10,000 individuals each — is then as much more intelligent than an average society of 10,000 humans as is a society of 10,000 humans than its average component human. Humans through language, writing and culture accumulate a store of solved problems for hundreds of generations. This observation applies more so for 6 billion human beings, who have moreover the heritage of infrastructures, cultures and sciences.

Mathematical concepts are more efficient, compressed, defined, more widely tested for longer periods of time, than concepts encoded into words. Societies test their own language over many generations, but all societies in all cultures have tested mathematics in millions of contexts repeatedly over hundreds of generations. Mathematics can be more precisely tested than words both through logical analysis and because the accuracy of mathematics used for modeling physical phenomena *can be measured*. Moreover, since the perceptions that inspire the creation of mathematical inference sets arise in nature, mathematical inference sets contain deep truths about the structure of the universe. Mathematics 'knows' things that individually we do not; to some extent individual mathematical reasoning involves data mining our collective store of mathematical inference sets. That being so, mathematical inference sets created by human societies should be able to predict aspects of phenomena that are implicit within them even when they are not apparent to human observers.

If mathematics distills collective intelligence, then that intelligence is vastly greater than that of any individual. Can the intelligence of mathematics, relative to an individual human being, be measured?

4 A formula for a society's problem solving rate

Suppose that the average individual problem solving rate in a society of n individuals is x problems per time unit, and the whole society's problem solving rate is X problems per time unit. Is there a function F such that $X = F(n)x$?

From the preceding sections conclude that networked brains are the source of solved problems that we acquire individually and use as a resource to either implement solutions to problems already solved by others or to solve novel problems. How is that information transmitted to us? Suppose that a population is isotropically networked so that each node in the network is equally likely to receive information, via the network, from any other node.

With the assumption of isotropy, if each individual can transmit information to μ nodes, each of those μ nodes receiving information can transmit in turn to μ nodes. Transmissions to all n individuals in a network would isotropically occur in $\log_\mu(n)$ generations of transmissions. From this infer that the function F in the equation $X = F(n)x$ is logarithmic. It is not sufficient to suppose that the function is logarithmic to use it to calculate rates of problem solving; we must know what number μ is. This differs from the usual situation of a problem solved by a logarithmic function, when it is a matter of indifference what the base of a logarithmic function is. Here discovering what the base of the logarithm must be is the problem.

In the isotropic situation, for $F(n) = \log_\mu(n)$, μ must be a parameter that results in F increasing the rate of information received by an individual who has problem solving rate x problems per time unit. This requires that the logarithm's base μ getting smaller increases F. μ must be proportional to the time required to transmit information to another; the less time it takes for a recipient to receive new information, the greater is the networking benefit to the recipient. Time is proportional to distance. A distance intrinsic to networks is the mean path length. It is moreover helpful that in the past 10 or 15 years the mean path length has been measured for various networks, including social networks and neuron networks, both of which are relevant to this analysis. This allows us to measure the rate of change per degree of freedom relative to the mean path length. The form of F is the same as the form of the logarithmic equation for entropy.

Claude Shannon in his seminal paper observed that information entropy is maximal when the probability of binary digits 0 and 1 is equal, an observation equivalent to supposing that the base of the logarithm that represents information entropy is the average probability. This is a particular instance of Jensen's inequality. Conclude from this that in general an isotropic distribution of information in a network — when the base of the logarithm (the average distance) between all nodes is the same — is maximally efficient: every recipient node has the same chance of receiving information from a transmitting node. By distance we mean the distance in connection steps. A step is popularly known as a degree of separation.

If a network's μ information transmitters isotropically scale by μ, conclude from this that $d\mu/dt = \mu$; the mean path length of the transmitting nodes is equal to e steps where e is the natural logarithm.

Since μ as the base of the logarithm acts as a scale factor, that suggests another connection to the natural logarithm. Consider a radial length that grows by a quantum length L itself proportional to a quantum of energy. The scale factor for length kL required to arrive at length $(k+1)L$ is $(k+1)/k = 1 + 1/k$. Now select an interval on the radial length, centered on kL, from $[-(1/2)k]L$ to $[+(1/2)k]L$. Choose the average scale factor $s = 1 + 1/k$ for all the increments of L that occur along that radial path. Then the cumulative scaling of the first length increment along that interval is about $(1 + 1/k)^k$. If L is small and k is large, then as k increases without bound, $(1 + 1/k)^k$ approaches the value of the natural logarithm. The assumption that each iteration of a radial length is constant is similar to assuming a constant speed of light. If this analogy is valid, it suggests a physical basis for the concept of counting (the number of iterations of energy quanta) and for the natural logarithm (the natural average scale factor for the linear propagation of light). Mathematical concepts that we construct by mining collective mathematical knowledge may reflect fundamental physical attributes of the universe.

What is the mean path length of information receivers compared to that of information transmitters? Rudolf Clausius, who developed much of the modern science of thermodynamics, solved this problem for an equivalent question. In his 1858 paper (p. 140 in Brush) he remarks about molecules in a gas:

> The mean lengths of path for the two cases (1) where the remaining molecules move with the same velocity as the one watched, and (2) where they are at rest, bear the proportion of $\frac{3}{4}$ to 1.

He gave a geometrical proof in 1860 (p. 434). Individuals randomly transmitting information are analogous to the randomly moving gas molecules, and the recipients of information are analogous to the gas molecules at rest. The transmitting system T and the recipient system R respectively have mean path lengths μ_T and μ_R. Since a network isotropically transmitting information has a mean path length equal to e, the recipient information network has a mean path length of $\mu_R = (4/3)e = (4/3)\mu_T \approx 3.624$. Another way of looking at this involves degrees of freedom. The amount of energy proportional to a length μ_T in T is proportional a length $(4/3)\mu_T = \mu_R$ in R. Accordingly, the same energy results in four degrees of freedom (*deg*) relative to μ_T in T and three *deg* in R, or a ratio of 4/3 *deg* : 1 *deg* for T compared to R. In 2001, Lawler, Schramm and Werner (Lawler 2001) found that the fractal dimension of Brownian motion is 4/3, which is consistent with Clausius's result.

One more step is required to find F. For an isotropically connected network we found that $F(n) = \log_\mu(n)$ where μ is the network's mean path length. Actual networks though are not isotropic; some people are closer, information-wise, than others. The parameter C, the clustering coefficient, is the average connectedness of nodes in a network to their adjacent nodes; for an isotropic network, $C = 1$. C has been measured for networks relevant to our problem. The general formula for the entropy (*deg*) of a network is therefore $F(n) = C \log_{\mu_R}(n)$; we will use it to quantify the problem solving capacity of a network compared to that of an individual within it.

Network	Nodes	Number of nodes	μ	C	F	Notes
C. elegans	neurons	282	2.65	0.28	1.62	1
Human Brain	neurons	10^{11}	2.49	0.53	14.71	2
1657 English	words	200,000	2.67	0.437	5.431	4, 5
1989 English	words	616,500	2.67	0.437	5.932	3, 4
1657 population	people	5,281,347	3.65	0.79	9.445	1, 7
1989 population	people	350,000,000	3.65	0.79	12.0	1, 6

Table 1: Calculations of F

Notes to Table 1
1. μ and C for C. elegans cited in, and for people based on an actors network study, Watts and Strogatz (1998).
2. The number of neurons: Nicholls, 2001, p. 480. μ and C: Achard, 2006.
3. The number of words: OED.
4. μ and C: Ferrer, 2001 based on about 3/4 of the million words appearing in the British National Corpus.
5. The number of words: EMEDD.
6. The number is estimated for 1989 by adding censuses: 1990 USA, 248.7 million; 1991 Canada 27,296,859; 1991 England 50,748,000; 1991 Australia, 16,850,540. These total 343,595,000 people.
7. The number of people in England: Table 7.8, following p. 207, for the year 1656, in Wrigley, 1989.

5 Calculating collective mathematical intelligence

This section's aim is to compare the mathematical problem solving capacity of society, X problems per time unit, to the mathematical problem solving capacity of an average human being, x problems per time unit. Two networks are involved. First there is the network of currently networked human brains. Since there is one brain per person, we are interested, for a given population Pop in the value of $F(Pop)$. The collective neuronal problem solving capacity of society at a given point of time is $F(Pop)x$, where x is the problem solving rate that an average individual is born with, what might be called average individual innate intelligence. The collective brain $F(Pop)x$ also has access to the contemporaneous network of inference sets which we here denote $F(Lex)$. Thus we have $X = F(Pop)F(Lex)x$.

The function F in effect measures the degrees of freedom of a network relative to its mean path length. The formula $X = F(Pop)F(Lex)x$ reveals the importance of knowledge (a store of solved problems) for a society and individuals within it. The problem solving capacity of a society is proportional to the degrees of freedom possessed by networked brains times the degrees of freedom of the society's store of inference sets. $F(Pop)F(Lex)$ must be multiplied by the rate x to give the rate X. Without problem solving devices such as brains, no new problems get solved. The more inference sets that a society has, the greater is its capacity X to solve problems but the more there is for individuals to learn. Efficient inference sets are easier and quicker to learn and to juggle than unwieldy ones, improving society's problem solving capacity. For example, a heliocentric solar system is conceptually simpler than a geocentric one.

If we confine ourselves to mathematical inference sets, we are immediately stymied in an attempt to estimate X because there is no easily identifiable collection of solved mathematical problems nor any obvious way to count solved mathematical problems. We need countable sets of collectively solved problems at different times that are large enough to allow us to estimate a problem solving rate X.

A language's lexicon is a large group of collectively solved problems that involves the whole of the society speaking that language. I propose to use the English lexicon as a countable collection of inference sets that allows us to calculate a rate representative of the rate of improvement in inference sets. Our calculations will use Table 1.

The measured value for μ for the English lexicon 2.67, in Table 1, is close to the value $e \approx 2.718$ which theory predicts for transmitting nodes. Similarly, the measured value for μ for a network of people receiving information is 3.65 in Table 1, very close to $(4/3)e \approx 3.624$ predicted for receiving nodes. The concurrence of theory and measurement supports the validity of using currently measured values of μ and C for $F(Pop)$ and $F(Lex)$ for earlier eras.

We can estimate X based on various inference sets. Having X will enable us to calculate x using $F(Pop)$ and $F(Lex)$ and to then compare x to X.

The Early Modern English Dictionaries Database (EMEDD) has about 200,000 word-entries for a period ending 1657. The OED has 616,500 words in 1989. The average English lexical growth rate in those 332 years based on this data is 3.39%. The economist William Nordhaus wrote an article on lighting to show that consumer price indexes do not capture technological advances. He found that the labor cost of light decreased from 41.5 hours of work for 1,000 lumen hours at 1,750 BCE to 0.00012 hours of work per 1,000 lumen hours at 1992 (his Table 1.4), an increase in efficiency of about 345,800 times. The average rate at which rate lighting efficiency increased is 3.41% per decade over a period of 3,742 years. Jim Oeppen and James Vaupel (2002) found that female longevity in Norway (one example in their article) increased from 47.9 years in 1841 to 77.32 years in 1970, which is 3.71% per decade, and that male longevity in Denmark (another example) increased from 43.11 years in 1840 to 56.69 years in 1919, which is 3.46% per decade, somewhat indicative but over short time periods. Manuel Eisner (2003) estimates the London homicide rate in 1278 at about 15 per 100,000 inhabitants (p. 84) compared to the English homicide rate in 1975 of 1.2 per 100,000 inhabitants (p. 99), a rate of decrease of 3.75% per decade.

Think of a society's average individual IQ as the society's average problem solving rate per capita. The rate of increase of average IQs is about 3.00 to 3.63 IQ points per decade (p. 113 and Table 1 at p. 180, Flynn 2007), consistent with the rates above. Infer, based on the closeness of the rates, that average IQs improve because inference sets improve.

Since the Nordhaus study covers the longest period of time of the above studies and is so close in value to the other sources of data, especially the large collection of solved English language problems represented by the English lexicon over 332 years, use as a long term average $X = 3.41\%$ per decade increase in problems solved.

Find x by solving $X = F(Pop)F(Lex)x$ for the period 1657 to 1989 using the Table 1 values. Use the measured $X = 3.391\%$ per decade for English lexical growth, which is close to 3.41% per decade based on the Nordhaus lighting study. For 1657 to 1989 average $F(Pop) = 5.682$ and average $F(Lex) = 10.72$, which gives an average $x = 5.56\%$ increase in solved problems per thousand years over that period. For 1657 to 1989, average X is 61.28 times average x. Put another way, society starting with 1,000 solved problems would add 34 solved problems in 10 years. It would take an individual deprived of society and its growing store of solved problems 1000 years to solve 5.6 problems.

There is a way to check the reasonableness of the estimated average $x = 5.56\%$ increase in solved problems per thousand years using glottochronology. Glottochronology, conceived by Morris Swadesh, estimates the date of origin of a mother language based on the rate of divergence between two daughter languages. Swadesh studied the historical usage of a set of 100 words likely to be relatively stable in most languages (such as I, you, he, we, here, there, etc.). About 1966 he estimated that cognates — words that sound similar in different daughter languages with a common mother language — diverge by a range of rates from 9% to 19% per thousand years (p. 280, Swadesh) for Indo-European, an average of about 14% per thousand years. For example, on these assumptions if the lexicons of two languages differ by 14% today, their common mother language existed 1,000 years earlier. Thirty-seven years after Swadesh's 1966 estimate, Russell Gray and Quentin Atkinson in 2003, with the use of a computer, estimated the beginning of the ancestral Indo-European language at 8,700 years before the present (the year 2003 then) which is about 6,700 BCE. To modernize Swadesh's estimate of the divergence rate of daughter languages by using the 2003 findings of Gray and Atkinson, take 7037/8700 times 14%, which equals 11.32% (the adjusted Swadesh rate), half of which is 5.66%, very close to the 5.56% found using F. Our calculation of x appears reasonable.

With the advantage of hindsight one can see why half the adjusted Swadesh rate must match the rate x that we found using F. Suppose a society divides into two daughter populations and that the original mother language has rate X_M. The two daughter languages (1 and 2) begin to diverge. The collective rate of X added solved problems per time unit for each daughter language is governed by $X = F(Pop)F(Lex)x$. Assume that on average, the size of the populations and lexicons for the two daughter languages and for the original mother population are soon the same, so that $F(Pop)$ and $F(Lex)$ are the same for all of them. Assume that the size of the mother language at the time when the daughter populations diverge is Lex and thereafter is unchanging. One thousand years later the size of the mother language is $Lex(1+0)$ (unchanged, as assumed) while a daughter language then has a lexicon of size $Lex(1 + x)$. Then the rate of change of a daughter language compared to the mother language is $[(1 + x)Lex - Lex]/Lex = x$. In other words, each daughter language diverges from the historical mother language at the rate x added solved problems per thousand years, so $2x$ should equal the adjusted Swadesh rate of divergence. (When I observed a whole number ratio between x and the adjusted Swadesh rate in June 2007, I felt that Mathematics must know the explanation though I did not, because the ratio was too precise to be a coincidence.)

That is why half the adjusted Swadesh rate, found using entirely different means, almost matches, and so corroborates, the average $x = 5.56\%$ we found using Table 1 values.

6 Observations and conclusions

Using Table 1, in 1989 $F(Pop)F(Lex) = 71.2$. How might we understand the implications of that? Since F is a logarithmic function, this 71.2 is similar to orders of magnitude except that F is based on mean path lengths which are smaller than 10. We may think of x as the average individual problem solving capacity before it is augmented by learning. In 1989, X, our estimate of the mathematical problem solving capacity of English speaking society based on an economic argument relating it to English lexical growth, was 71.2 times as great as the average innate human problem solving capacity. Taking the population of the whole world into account, and the fact that it is now 2015, X is probably modestly larger, F being logarithmic.

Here is a suggestion for understanding how much we owe to society when we use mathematical methods and concepts. Using Table 1, F for the neurons of a human brain is 9.08 times bigger than F for the neurons of the worm C. elegans. Suppose F is also a rough measure of problem solving capacity when applied to networked neurons. The problem solving capacity of society is about 71.2 times the innate problem solving capacity of an average individual. The ratio of society's problem solving capacity compared to that of the average individual is 71.2/9.08 = 7.8 times greater than the ratio of the neuronal capacity of a human being compared to that of the worm C. elegans. Is it any wonder that we should regard mathematics as ineffable? Relative to an individual, society's problem solving capacity using mathematics is almost omniscient.

To end this essay, a quote from a 1946 Hollywood movie, A Guy Named Joe (p. 58, Cairn), seems apt. For "learned to fly" in the first line of the quote substitute "learned how to use mathematics" (and learning inference sets in general), and similarly for the other references to flying.

> ... the General interrupted him. "You're not under the impression that you learned to fly all by yourself, are you?" He asked quietly. Peter glared. "Well, if I didn't, sir, I'd like to know who helped me." "Well, I guess I can tell you that. .... You were helped by every man, since the beginning of time, who dreamt of wearing wings. By pioneers who flew pieces of wire and pasteboard ... long before you were born. By every pilot who ever crashed into the ground in order that others would stay up in the sky."

References

[1] Achard, S., Salvador, R., Whitcher, B., Suckling, J. & Bullmore, E. (2006) A Resilient, Low-Frequency, Small-World Human Brain Functional Network with Highly Connected Association Cortical Hubs. The J. of Neuroscience, 26(1), pp. 63-72.

[2] Cairn, J. (1946). A Guy Named Joe. MGM Film Classics Series, London.

[3] Clark, J. B. (1899). The Distribution of Wealth: A Theory of Wages, Interest and Profits. MacMillan.

[4] Clausius, R., 1858. Ueber die Art der Bewegung, welche wir Wärme nennen. Annalen der Physik 101, pp. 353–380; English translation, The Nature of the Motion which we call Heat, in S. G. Brush, 1948, Kinetic Theory, Pergamon Press, Vol. 1, 135.

[5] Clausius, R., 1860. On the Dynamical Theory of Gases, Philos. Mag. J. of Sci, Vol. 19, 4^{th} Series, Taylor and Francis.

[6] Davis, P. J., Hersh, R. & Marchisotto, E. A. (1995). The Mathematical Experience - Study Edition. Boston: Birkhauser.

[7] EMEE (1999). www.chass.utoronto.ca/english/emed/#dic as of October 15, 1999.

[8] Eisner, M. (2003). Long-Term Historical Trends in Violent Crime. Crime and Justice; A Review of Research, 30, 83-142.

[9] Ferrer i Cancho, R. & Solé, R. V. (2001). The Small-World of Human Language. Proceedings of the Royal Society of London, B 268, 2261-2266.

[10] Flynn, J. R. (2007). What is Intelligence? Cambridge University Press.

[11] Gray, R. D. & Atkinson, Q. D. (2003). Language-Tree Divergence Times Support the Anatolian Theory of Indo-European Origin. Nature 426, pp. 435-439.

[12] Hamming, R. W. (1980). The Unreasonable Effectiveness of Mathematics. Am. Math. Mon. 87(2).

[13] Hayek, F. A. (1948). The Use of Knowledge in Society, in Individualism and Economic Order. University of Chicago Press.

[14] Heath, T. (1921). A History of Greek Mathematics. Oxford.

[15] Jevons, W. S. (1879). The Theory of Political Economy, Second edition. MacMillan.

[16] Lawler, G. F., Schramm, O., Werner, W., (2001). The Dimension of the Planar Brownian Frontier Is 4/3, Math. Res. Lett. 8, 401.

[17] Nicholls, J. G., Martin, A. R., Wallace, B. G. & Fuchs, P. A. (2001). From Neuron to Brain. Sinauer.

[18] Nordhaus, W. D. (1997). Do real-output and real-wage measures capture reality? The history of lighting suggests not. Cowles Foundation Paper No. 957, and in Gordon, R. J. & Bresnahan, T. F. (ed.) The Economics of Goods, U. of Chicago Press.

[19] Oeppen, J. & Vaupel, J. W. (2002). Broken Limits to Life Expectancy. Science 296, 1029.

[20] Sarukkai, S. (2005). Revisiting the 'unreasonable effectiveness' of mathematics. Curr. Sci. India 88(3).

[21] Strogatz, S. (2003). Sync. New York: Hyperion.

[22] Swadesh, M. (1971). The Origin and Diversification of Language. Aldine-Atherton.

[23] Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. Nature 393, 440.

[24] Wigner, E. (1960). The Unreasonable Effectiveness of Mathematics in the Natural Sciences. Comm. Pure Appl. Math. 13(1).

[25] Wrigley, E. A., Schofield, R. & Lee, R. D. (1989). The population history of England, 1541-1871: a reconstruction. Cambridge University Press.

[26] Census - Australia—1991 Census of Population and Housing. Australian Bureau of Stat, 1993, Cat. No. 2710.0.

[27] Population and Dwelling Counts, Canada, Provinces and Territories, 1991 and 1996 Censuses. Stat. Can., Cat. No. 93-357-XPB.

[28] Revised population estimates England & Wales 1991-2000. Nat. Stat., 2003.

[29] USA Age 2000. nationalatlas.gov, 2000, Census 2000 Brief Series.