Pushing the Complexity Barrier: Diminishing Returns in the Sciences

Claudius Gros

Institute for Theoretical Physics
J. W. Goethe University
Frankfurt, Germany

Are the sciences not advancing at an ever increasing speed? This popular perspective is contrasted with the view that scientific research is actually closing in on complexity barriers and that, as a consequence, science funding actually sees diminishing returns, at least in established fields. In order to stimulate a larger discussion, two exemplary cases are investigated: the linear increase in human life expectancy over the past 170 years and advances in the reliability of numerical short- and medium-term weather predictions during the past 50 years. It is argued that the outcome of science and technology funding in terms of measurable results is a highly sublinear function of the amount of resources committed. Supporting a range of small to medium research projects instead of a few large ones will be, as a corollary, a more efficient use of resources for science funding agencies.

1. Measuring Scientific Progress

There is a curious dichotomy in our current science and technology (S&T) landscape. On one side, we see advances on scales as never before in human history. There is, on the other side, a growing sentiment among researchers that progress in science is becoming increasingly harder to achieve. This sentiment is based in part on anecdotal evidence that is continuously reinforced by new insights. On the anecdotal side, there is the phenomenon that the requirements for a typical Ph.D. thesis in the natural sciences have increased dramatically over the past 50 years. It is well known that nowadays it takes much longer for a young scientist to reach the forefront of research.

The notion that scientific research needs to deal with rising levels of complexity is especially evident when studying the realm of life. An example of new insights bolstering this notion are the results of the ENCODE project [1], showing that our genome not only contains 21 000 protein encoding genes, but also has up to 4 million regulatory switches where transcription factors could bind, besides a myriad of other regulatory sequences. Is it possible to quantify this notion of a rising complexity level? This is the central topic of our investigation and it involves the quest to actually measure the pace of scientific progress.

Scientific progress is notoriously hard to measure. It does not really make sense to quantify advances in fundamental research; a single publication leading to a paradigm shift may be invaluable. However, the vast majority of scientific research efforts are directed toward achieving incremental progress, and are not of a foundational character. Hence, it is worthwhile to ask how taxpayers' money allocated for public science funding could be spent most efficiently.

In order to make a first inroad, we investigate two large-scale endeavors of humanity. The first case study concerns the long-term impact of research and investments in medicine and healthcare on life expectancy over the past 170 years. We ask the somewhat antipolar question: why did the average life expectancy rise so slowly? The second example regards advances in the reliability of short- and medium-term numerical weather forecasts since the 1950s. Weather dynamics have potentially chaotic regimes and the pace of progress in predictive meteorology may be limited by a complexity barrier resulting from systemic difficulties in predicting chaotic dynamical systems.

We find that measurable progress in S&T is a highly sublinear function of the invested resources, reflecting the law of diminishing returns that is well studied in economical contexts [2]. Some scientific insights can be achieved only through large collaborative projects, like the search for the Higgs boson. However, our results show that small scientific endeavors do generically offer a higher potential for returns in terms of results per allocated funding.

Many natural systems investigated in the sciences are complex dynamical systems [3]. The brain in the neurosciences and the human genome in bioinformatics are examples from the realm of life. Short- and long-term weather and climate evolution, many-body systems in condensed matter physics, and elementary particle condensates are examples from the realm of physics. Complex systems are both difficult to understand and investigate on a conceptual basis, as well as to model and simulate numerically. These two difficulties impinge the pace of progress when investigating complex biological or physical systems. We find the notion of a malleable complexity barrier to be a good visualization for the challenges confronting scientists today.

2. Record Life Expectancy

Human life expectancy has seen a dramatic rise since the middle of the eighteenth century. On a global level, the "record life expectancy" is commonly considered as the life expectancy at birth of the country having the highest life expectancy worldwide. The record life expectancy has seen a strikingly linear growth for about 150 years, as shown in Figure 1. It has been repeatedly predicted that this steady increase of human life expectancy would need to level off at a certain point, invoking biological limits. All these predicted limits have been broken hitherto without exception [4, 6]. This spectacular steady growth of the record life expectancy raises a series of interesting points.

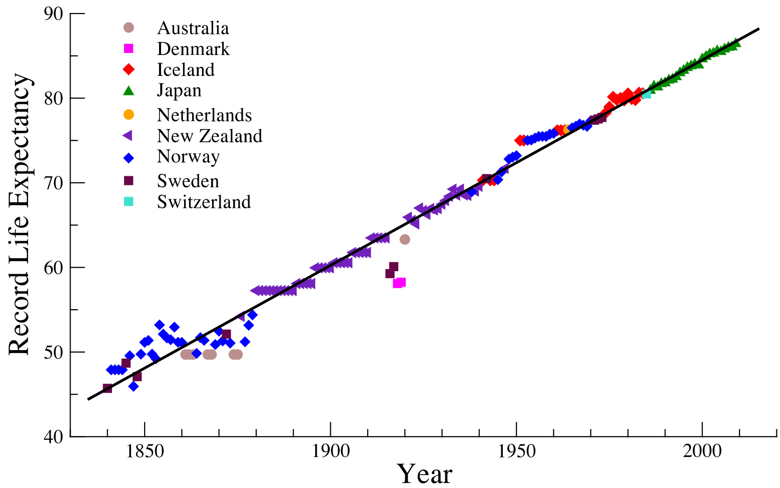

Figure 1. Record female life expectancy [4, 5]. This plot shows the life expectancy of the country with the highest average female life expectancy in a given year, for all calendar years. The line is a least square linear regression with the increase in record life expectancy averaging 2.4 years for 10 calendar years.

Advances in healthcare and medicine will lead generically to raising life expectancy. However, it remains unclear which forces determine the magnitude of the observed rate of 2.4 years per every 10 calendar years and how the observed growth rate depends on the overall amount of resources devoted [7]. On one hand, breakthroughs in research have been postulated to boost human lifespans rapidly [8]. On the other hand, if a putative natural limit exists [4, 9], it should lead to a gradual leveling off. We have yet no definite answers to these fundamental questions.

Longevity is a central issue in our culture and major efforts and resources are devoted by our societies toward increasing health levels and lifespan. Figure 1 demonstrates that returns on investments have dramatically decreased during the past 150 years. The initial growth in life expectancy resulted from simple hygiene measures, followed by progress in immunology and antibiotics research. Lately, massive investments in pharmacology, technical medicine, and bioinformatics have been necessary to keep up the steady linear advance in life expectancy. Relative progress has actually decreased in spite of these massive efforts; a linear increase relative to a base of 45 years is twice as large as a linear increase (with the same slope) relative to a base of 90 years.

Investments in medicine and healthcare have seen a roughly exponential increase during the past half a century [12, 13]. The driving forces behind these ever-rising costs are debated and could be rooted either in the desire to increase health and well-being quite generally or, more directly, in the quest to postpone death as far as possible. It has been argued, in this context, that the economic rationale behind the ever-rising levels of healthcare spending lies in the fact that humans attribute an income elasticity well above unity to improvements of life expectancy, which seems to have psychologically a nondeclining marginal utility [14]. This argument indicates that the average life expectancy is indeed a valid yardstick for measuring the overall advancement in health and medicine. The quantitative efficiency of S&T research efforts in the healthcare sector is hence, as measured by the observed extension of human life expectancy, a highly sublinear function of resources devoted. Returns on investment are seen to diminish rapidly in medicine and healthcare.

3. Weather Forecast Reliability

The dynamics of weather and climate for medium- to long-term time scales is known to contain chaotic components [15]. Indeed, the Lorenz model [16], one of the central models in the theory of chaotic and complex systems, is a hydrodynamic convection model. Medium- to long-term weather forecasting is hence considered a challenge and massive investments in modeling, data acquisition, and computational infrastructure have been made in order to achieve improved forecast performances.

A range of numerical forecasting skill scores for short- to medium-term weather prediction are evaluated continuously. This is done in order to track the quality of daily weather predictions by national and international weather and climate agencies. Figure 2 shows the historical evolution of two distinct prediction reliability measures. The first is the 1 through 7 days 500 hPa forecast correlation coefficients of the Germany Weather Service (DWD) [11, 17]. A value close to unity signals optimal forecasting, while values below 60% correspond to essentially useless predictions. The introduction of a new model in 1990 is reflected in the data. Also given in Figure 2, for a longer-range perspective (1955–1991), is the S1 36 hour skill score of the United States National Oceanic and Atmospheric Administration (NOAA) [10, 18], which we normalized to the interval [0, 1]. The S1 skill score contains gradients so it is qualitatively different from the 500 hPa correlation coefficient [18], so a direct comparison of the absolute values is not meaningful for these two weather forecasting reliability measures.

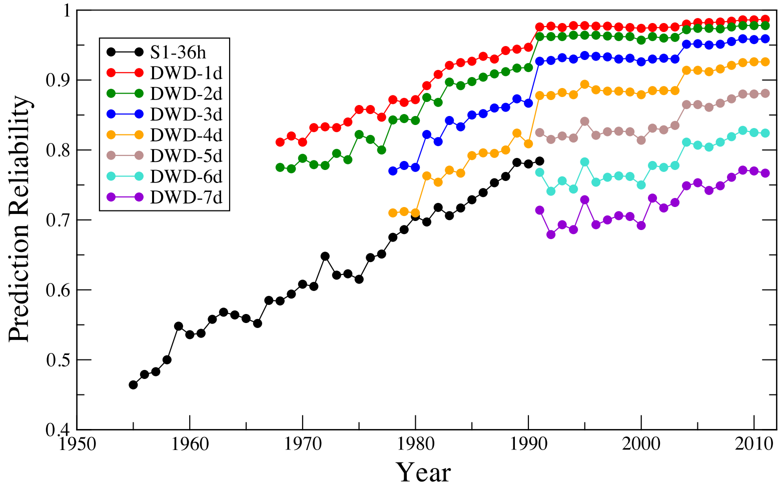

Figure 2. Two distinct weather forecast reliability measures are shown, the 500 MB 36 hour S1 score of the NOAA [10] (black data), and the 1 through 7 days 500 hPa correlation coefficient of the DWD [11] (color data). Note that these two reliability scores are differently defined and cannot be compared directly on a quantitative basis.

One of the key ingredients for numerical weather prediction, besides modeling and data acquisition [18], is computing power. Available computing power has seen an exponential growth over the past 60 years following Moore's law [19, 20] with a doubling period of about 1.5 years. The resulting advance in computational power has been about 10^{10} in 50 years. The computational facilities employed for numerical weather forecasting have seen corresponding increases [18], contributing to the observed improvements in weather forecasting skills.

In order to estimate the scaling between computational resources and forecasting skills quantitatively, Figure 3 shows the standardized forecasting error (1 – reliability) corresponding to the remaining difference to optimal forecasting. Because the S1 skill score data is so distant from optimality, we focus on the DWD data for a systematic analysis. In Figure 3 we have replotted the DWD data as $-\log(1 - \text{reliability})$, using a three-year trailing average as a smoothing procedure. In order to compare the growth rates of prediction accuracy for different forecasting timescales, we analyze the data presented in Figure 3 using least square regressions. They fit the data reasonably well, indicating that the long-term growth of prediction accuracy is roughly exponential.

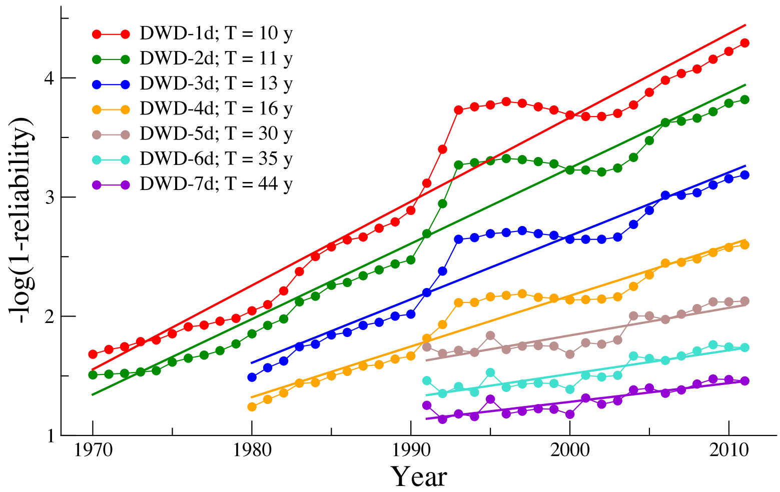

Figure 3. Linear–log plot of the DWD data (three years trailing average) of Figure 2. The lines are least square regressions; the inverse slope in terms of the number of years T needed to double the relative precision is given in the legend.

The time needed to double the relative accuracy, that is, to halve the forecasting error (1 – reliability), grows systematically with an increasing forecasting timespan (see the legend of Figure 3). The accuracy of one-day weather predictions has doubled historically roughly every 10 years. On the other hand, about 40 years seem to be necessary for improving the reliability of seven-day weather forecasts by a corresponding factor. Out of these results we conclude the following.

First, the quantitative progress in weather prediction accuracy is a highly sublinear function of committed computing resources. The accuracy scales roughly, with respect to the power P_c of the computer facilities employed, as $\left(P_c^{1.5}\right)^{1/10} = P_c^{0.15}$ for one-day forecasting and as $\left(P_c^{1.5}\right)^{1/40} = P_c^{0.0375}$ for seven-day predictions. The reliability of the skill scores is dependent additionally on advances in modeling and data acquisition; however, the respective functional relations of these dependencies are beyond the scope of the present discussion.

Second, forecasting becomes more difficult with increasing prediction periods. Indeed, it has been suggested that it may essentially be impossible to achieve useful forecasting reliabilities for 14 to 21 days in advance [21], at least with economically justifiable amounts of resources. This can be seen by plotting the weather forecasting scores as a function of the forecasting period, as shown in Figure 4, where we plot the T255L40 reliability score of the European Center for Medium-Range Weather Forecasts (ECMWF) from 2001 [21]. Also shown in Figure 4 are visual guides in the form of two-parameter least-square fits of the functional form

$$1 - \frac{t}{t + a\exp(-bt)}, \tag{1}$$

where $t = 1, 2, \ldots$ days, with adjustable parameters a and b. This functional form captures the notion that it becomes progressively more difficult to achieve reliable forecasting with increasing prediction periods. The functional form of this increase in complexity has been assumed here to be exponential.

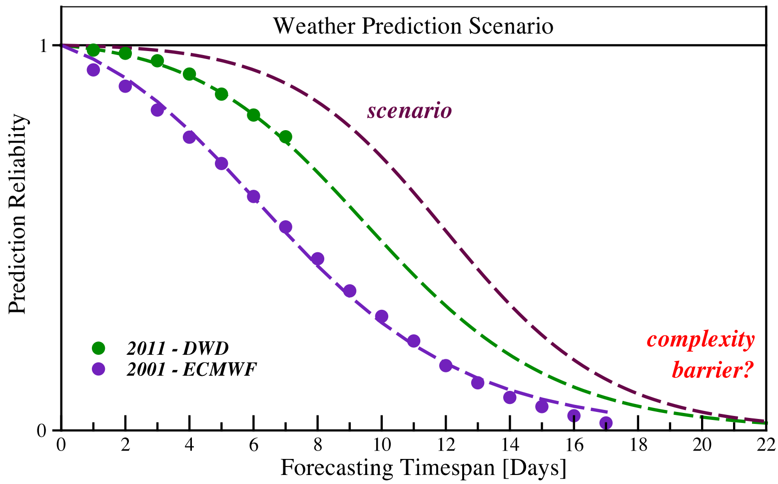

Figure 4. Forecasting reliability of the ECMWF T255L40 score from 2001 [21] (violet data), renormalized to [0, 1], and of the 2011 DWD 500 hPa score (see Figure 2, green data). The dashed lines are two-parameter fits using equation (1), intended as visual guides. The brown dashed line represents a putative scenario for reliability scores achievable with further advancements.

Generalizing these two observations, we may relate the measured S&T progress to the amount of committed resources via the scaling relation

$$\text{progress} \propto (\text{resources})^{\alpha}, \tag{2}$$

with a sublinear scaling exponent $\alpha < 1$. Our results, as shown in Figures 3 and 4, indicate that the scaling exponent α rapidly drops toward zero with increasing complexity of the task to be solved. We propose to use the term *complexity barrier* for this phenomenon. For the case of extending human life expectancy, which is rising linearly, invested resources have increased roughly exponentially [12, 13], leading to a logarithmic relation, progress $\propto$ log(resources). A log-relation corresponds to a vanishing scaling exponent $\alpha \to 0$, indicating that increasing the average human life expectancy is a task of very high complexity.

4. Entering the Human Factor

The complexity barrier present in the context of short- to medium-term weather forecasting is not hard. Progress is achievable when committing increasingly larger amounts of resources. The same holds for the complexity barrier present in the aging problem. Extending the average human life expectancy has been possible for the past 170 years by devoting strongly increasing amounts of resources to medicine and healthcare. The magnitude of the resources committed has been growing not only in absolute terms, but also as a fraction of gross national products (GNPs). The rationale for the underlying collective decision of resource distribution is to be attributed to the human factor; progress in extending the human life span is highly valued.

We have discussed here only two examples, but believe that progress in science, whenever it can be measured on a quantitative basis, is quite generically a strongly sublinear function $f(x)$ of the amount x of resources devoted. Sublinear dependencies are concave and for any concave function $f(x)$ the total return $\sum_i f(x_i)$ is larger when splitting the total amount of resources x into a series of subpackages of smaller sizes x_i,

$$f(x) < \sum_i f(x_i), \quad \sum_i x_i = x.$$

For funding agencies this indicates that a substantially more efficient use of available resources is to prioritize small to medium projects, whenever feasible.

On a last note, it is interesting to speculate whether the human factor influences the pace of progress in S&T in an even more direct way. The human brain is well known to discount incoming information streams logarithmically, a relation known as the Weber–Fechner law [22–24]. This exponential data compression is necessary in order not to drown in the daily flood of sensory impressions. It has been observed recently that these neuropsychological constraints shape the statistics of global human data production in terms of data files that are publicly available from the internet [25]. The human factor is hence in evidence, at least in this particular aspect of human S&T efforts, in the statistics of global public data generation. It is hence conceivable, as a matter of principle, that the pace of progress close to a complexity barrier is not only influenced by the overall amount of financial and physical resources committed, but also more directly by the neuropsychology of human thought processes.

Acknowledgments

I would like to thank Ulrich Achatz, George Craig, Ulrich Damrath, Peter Lynch, Detlev Majewski, Peter Nevir, and Johannes Quaas for support collecting the meteorological skill scores, as well as Roland Rau and James Vaupel for the life expectancy data.

References

[1] E. Pennisi, "ENCODE Project Writes Eulogy for Junk DNA," *Science*, **337**(6099), 2012 pp. 1159–1161. doi:10.1126/science.337.6099.1159.

[2] A. Marshall, *Principles of Economics*, 8th ed., New York: Cosimo Classics, 2009.

[3] C. Gros, *Complex and Adaptive Dynamical Systems, A Primer*, 2nd ed., New York: Springer, 2010.

[4] J. Oeppen and J. W. Vaupel "Broken Limits to Life Expectancy," *Science*, **296**(5570), 2002 pp. 1029–1031. doi:10.1126/science.1069675.

[5] Data since 2001 courtesy R. Rau and J. Vaupel. Human Mortality Database. (Oct 29, 2012) http://www.mortality.org.

[6] K. M. White, "Longevity Advances in High-Income Countries, 1955–96," *Population and Development Review*, **28**(1), 2002 pp. 59–76. doi:10.1111/j.1728-4457.2002.00059.x.

[7] S. Tuljapurkar, N. Li, and C. Boe, "A Universal Pattern of Mortality Decline in the G7 Countries," *Nature*, **405**, 2000 pp. 789–792. doi:10.1038/35015561.

[8] A. D. N. J. de Grey, "The Foreseeability of Real Anti-aging Medicine: Focusing the Debate," *Experimental Gerontology*, **38**(9), 2003 pp. 927–934. doi:10.1016/S0531-5565(03)00155-4.

[9] J. Vijg and Y. Suh, "Genetics of Longevity and Aging," *Annual Review of Medicine*, **56**, 2005 pp. 193–212. doi:10.1146/annurev.med.56.082103.104617.

[10] R. Y. Hirano. "The National Meteorological Center's Historical 36-(30-) Hour S1 Score Record." National Oceanic and Atmospheric Administration Office Note 389, 1992. http://www.ncep.noaa.gov/officenotes/NOAA-NPM-NCEPON-0005/014089CB.pdf.

[11] Data courtesy German Meteorological Service (DWD), Ulrich Damrath.

[12] G. F. Anderson, J. Hurst, P. S. Hussey, and M. Jee-Hughes, "Health Spending and Outcomes: Trends in OECD Countries, 1960–1998," *Health Affairs*, **19**(3), 2000 pp. 150–157. doi:10.1377/hlthaff.19.3.150.

[13] A. Sisko, C. Truffer, S. Smith, S. Keehan, J. Cylus, J. A. Poisal, M. K. Clemens, and J. Lizonitz, "Health Spending Projections through 2018: Recession Effects Add Uncertainty to the Outlook," *Health Affairs*, **28**(2), 2009 pp. w346–w357. doi:10.1377/hlthaff.28.2.w346.

[14] R. E. Hall and C. I. Jones, "The Value of Life and the Rise in Health Spending," *The Quarterly Journal of Economics*, **122**(1), 2007 pp. 39–72. doi:10.1162/qjec.122.1.39.

[15] J. Shukla, "Predictability in the Midst of Chaos: A Scientific Basis for Climate Forecasting," *Science*, **282**(5389), 1998 pp. 728–731. doi:10.1126/science.282.5389.728.

[16] E. N. Lorenz, "Deterministic Nonperiodic Flow," *Journal of the Atmospheric Sciences*, **20**(2), 1963 pp. 130–141. doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

[17] A. H. Murphy and E. S. Epstein, "Skill Scores and Correlation Coefficients in Model Verification," *Monthly Weather Review*, **117**(3), 1989 pp. 572–581. doi:10.1175/1520-0493(1989)117<0572:SSACCI>2.0.CO;2.

[18] P. Lynch, "The Origins of Weather Prediction and Climate Modeling," *Journal of Computational Physics*, **227**(7), 2008 pp. 3431–3444. doi:10.1016/j.jcp.2007.02.034.

[19] G. E. Moore, "Cramming More Components onto Integrated Circuits," *Proceedings of the IEEE*, **86**(1), 1998 pp. 82–85. doi:10.1109/JPROC.1998.658762.

[20] S. E. Thompson and S. Parthasarathy, "Moore's Law: The Future of Si Microelectronics," *Materials Today*, **9**(6), 2006 pp. 20–25. doi:10.1016/S1369-7021(06)71539-5.

[21] A. J. Simmons and A. Hollingsworth, "Some Aspects of the Improvement in Skill of Numerical Weather Prediction," *Quarterly Journal of the Royal Meteorological Society*, **128**(580), 2002 pp. 647–677. doi:10.1256/003590002321042135.

[22] S. Hecht, "The Visual Discrimination of Intensity and the Weber–Fechner Law," *Journal of General Physiology*, **7**(2), 1924 pp. 235–267. doi:10.1085/jgp.7.2.235.

[23] A. Nieder and E. K. Miller, "Coding of Cognitive Magnitude," *Neuron*, **37**(1), 2003 pp. 149–157. doi:10.1016/S0896-6273(02)01144-3.

[24] S. Dehaene, "The Neural Basis of the Weber–Fechner Law: A Logarithmic Mental Number Line," *Trends in Cognitive Sciences*, **7**(4), 2003 pp. 145–147. doi:10.1016/S1364-6613(03)00055-X.

[25] C. Gros, G. Kaczor, and D. Marković, "Neuropsychological Constraints to Human Data Production on a Global Scale," *European Physical Journal B: Condensed Matter and Complex Systems*, **85**(1), 2012 p. 28. doi:10.1140/epjb/e2011-20581-3.