The Price of Big Science:
Saturation or Abundance in Scientific Publishing?

Caroline S. Wagner, Dae Joong Kim
The Price of Big Science: Saturation or Abundance in Scientific Publishing?
Caroline S. Wagner, Ph.D., A Dae Joong Kim, B

Science policymaking is facing a rapidly changing landscape. Rapid growth and globalization of science are complicated by the proliferation of venues for publications, which continue to grow in number at an exponential rate. The growth rate is nullifying the hypothesis about its trajectory put forth by Derek de Solla Price (1961 and 1963); he suggested that science would reach a saturation point. In fact, the current system is proliferating, not just in numbers of published articles but also in the geographic location where knowledge is produced and in the types of venues for output (such as open source). The knowledge production system shares features with complex systems, so we propose a complex systems model to test the hypothesis. The model is designed along a stock and flow relationship between knowledge creation and obsolescence that tracks closely with actual numbers. The model further suggests that the publication system will continue to see exponential growth, and with this, may have experienced a phase shift from operating under conditions of scarcity to one of abundance. Abundant systems are characterized by openness, collaboration, and sharing— all features seen in contemporary science. Policymakers may need to shift policy toward scanning and integrating abundant knowledge to account for its proliferation and distribution across the growing knowledge landscape.

Keywords: science policy; knowledge production; complex systems modeling; collaboration

Introduction

The output of public investment in science and technology is contained in published articles, notes, and letters in scientific journals. These communications can be assumed to be part of a communications system with dynamics. The system has features of complex systems in that it is nonlinear, it has multiple agents competing for resources, and the resulting organization is emergent. Policymakers would like to understand, predict, measure, and control the processes in order to ensure that benefits accrue to intended beneficiaries. The complex system of scientific publication remains poorly understood, however. This article addresses this gap.

The eco-system of scientific communications has a number of levels, including

A Dr. Caroline S. Wagner holds the Ambassador Milton A. and Roslyn Z. Wolf Chair in International Affairs at the John Glenn School of Public Affairs, The Ohio State University, Columbus, OH. She earned her doctorate from the University of Amsterdam in Science and Technology Dynamics, her Master of Arts degree in Science, Technology and Public Policy from George Washington University, and a Bachelor of Arts from Trinity College, Washington, DC.

B Dae Joong Kim is a lecturer at Dongguk University in the Republic of Korea. Mr. Kim is a specialist in complex systems and public policy.
The Price of Big Science

Researchers operating largely within academic institutions. Researchers span institutional boundaries to self-organize into disciplines. Disciplinary organization is stabilized within journals reporting on activities within the field. Funding agencies also largely organize along disciplinary lines. Funding is based upon past publication performance and excellence as measured by citations to previous work in high impact journals.

Various attempts have been made to measure the scope and trajectory of scientific publishing, although not in ways that reveal systemic features. The literature on the growth of science has counted the number of articles and the proliferation of journals and other venues (such as online journals) where science is published. The rise of new venues of publication—such as arXiv, the Public Library of Science (PLoS), and Scientific Commons—complicates the counting and categorization processes. On the one hand, it shows vitality within scientific research, and on the other, it challenges those policymakers who rely on databases such as Web of Science and Scopus, which, for all their strengths, are limited in their ability to represent the scope of scientific output (Larsen and Von Ins 2010; Stevan Harnad 1997; Christine Borgman, Wallis, and Enyedy 2007). The databases are highly skewed toward reporting elite science rather than all of the scientific outputs.

Price’s hypothesis of growth of scientific knowledge

The rapid growth of science is not new: even in the seventeenth century, scholars bemoaned the problem of keeping up with the rush of publications. In 1613, Barnaby Rich wrote: “One of the diseases of this age is the multiplicity of books! They doth so overcharge the world that it is not able to digest the abundance of idle matter that is every day hatched and brought forth into the world.” In an attempt to characterize contemporary science, most scholars turn to the seminal work of Derek de Solla Price, who began his studies of science in the early 1960s. In this case, we will test Price’s hypothesis—tied to the “Big Science” model—of the saturation point of scientific publishing. In 1961, Price pointed out that scientific output had been growing at an exponential rate for three centuries. He further noted that the laws of physics would suggest that nothing can continue to grow indefinitely along an exponential growth curve. With this in mind, Price proposed a saturation point, shown in Figure 1, where he hypothesized that the growth trend of scientific publication would follow an s-curve in that knowledge would reach a saturation point. He anticipated that this point would occur within a generation from his writing (which would be contemporaneous with this article) and then would level off. The growth of scientific knowledge would reach a point where its growth would shift from exponential to a steady state of production.

To support his argument, Price analyzed the total number of papers published in scientific fields since 1740. Based on the historical trend of the published papers, the finite nature of growth, and the known numbers at the time of calculation, Price expected that science would reach a point where exponential growth was no longer possible. Price’s thesis and associated literature has been well reviewed by many scholars, including Jinha (2010), Lariviere, Archambault, and Gingras (2008) and Fernandez-Cano, Torralbo, and Vallejo (2004). Lariviere, Archambault, and Gingras (2008) reviewed the literature on the lifecycle of scientific literature. They calculated Price’s Index (1986) of knowledge obsolescence for
Figure 1. Price's knowledge saturation model
Source: Price (1960)
100 years of publications in Web of Science based on citations. They show that even with the rapid increase in output, knowledge is not becoming obsolete more rapidly. Moreover, the average age of cited literature has held fairly stable at eight years for nearly 50 years (p. 295). They calculate Price's Index (1963) of obsolescence and found that the Internet may be enabling greater access to older literature, which in turn is being cited. They infer that this access may be slowing the rate of obsolescence.

Similarly, Jinha (2010) found that scientific output has kept growing exponentially from the time that Price conducted his calculations. Jinha (2010) investigated annual global research output patterns based on the amount of accumulated scientific articles between 1726 and 2009 and found that the growth of publications at the article level is growing along the pure exponential growth curve in Price's model in Figure 1, not along Price's saturation curve¹ (see appendix). Table 1 shows supporting data.

As Price himself noted in later works (1986), electronic formats and storage capabilities have presented to science the enormous potential to manage and develop databases in formats that did not exist when he first proposed the saturation hypothesis. The advent of the Internet provides the ability to create new electronic journals and new forms of data storage. Whether directly related to the Internet or to other dynamics or some combination in between, scientific knowledge in terms of articles as well as venues (journals and online sites) is growing at a significant rate without having reached Price's anticipated saturation point. Indeed, Larsen and Von Ins (2010)² find that traditional scientific publishing (publication in peer-reviewed journals) has not given way to online publishing, and is still increasing, although with some notable differences between fields in growth rates.

**Why is scientific publishing not “saturated”?**

We explored the question of why scientific knowledge has not reached the saturation point Price anticipated. Perhaps the most obvious reason is the emergence of electronic storage and diffusion technologies. In 1960, paper libraries imposed physical constraints that limited the ability of scientists to access and build upon earlier work. These limitations have been overcome or at least mitigated³ with the advent of digital abstracting and publishing, including the creation of online resources and the rise of open access journals; the access to archives of formerly unpublished, primary source materials now available online; and massive electronic storage capacity (with anticipated access to full text still somewhat limited). Moreover, it would not have been obvious to Price that, by the twenty-first century, developing countries would rapidly join the ranks of scientific research, and further, that their researchers would seek publication outlets at the same rate as those in advanced countries (Wagner and Wong, 2011).

---

¹ Among those scholars who have tried to count the numbers of articles and journals being published in one year or accumulative over time, the estimates range fairly widely depending upon the assumptions about the boundaries of scientific scholarship, as well as the database from which the analysts draw.

² Larsen and Von Ins analyzed available data between 1907 and 2007 from a number of literature databases, including Science Citation Index (SCI) and Social Sciences Citation Index (SSCI).

³ Some say that the continuing practice of having scientific articles behind the ‘walls’ of subscription-based journals limits access for many potential readers, see Harnad (2007).
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Indicator</th>
<th>Value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial knowledge</td>
<td>Number of scientific articles published in 1990</td>
<td>842,531</td>
<td>Jinha (2011)</td>
</tr>
<tr>
<td>(Fractional) creation rate</td>
<td>Average growth rate of published articles</td>
<td>0.325</td>
<td>Mabe (2003)</td>
</tr>
<tr>
<td>Carrying capacity</td>
<td>Average number of scientific articles to be published in journals per year</td>
<td>22.978</td>
<td>Vinkler (2004)</td>
</tr>
</tbody>
</table>

**Table 1.** Data underlying assumptions for revisiting Price’s model

**Figure 2.** Research outputs in numbers of articles, 1726–2009
The quality of the expanding output is a significant question that has not gone without comment, even by Derek Price himself, as well as philosophers of science, Robert Merton, Thomas Kuhn, and Karl Popper. Indeed, Popper (1976) viewed the growth of scientific output as a deficiency of science, since, in his view, the “growth of knowledge… is not a repetitive or cumulative process, but one of error elimination…” (p.144) and therefore should not increase in volume, but decrease as errors are eliminated from the understanding of the natural world. Price also wrestled with the quality question, noting that perhaps it would be better to take a different approach to science: “One may study the growth of only important discoveries, inventions, and scientific laws, rather than all such things…” (1961, p. 32), thus reducing the overall output of science.

Indeed, several philosophers of science have suggested that science should be conducted by an elite group of highly talented researchers, rather than a broad group. Bradford’s Law (Garfield 1971) informed the earliest efforts by Francis Narin, Eugene Garfield, and Henry Small to identify high quality science and are based upon the premise that a small percentage of output would represent the highest quality work; thus a limited dataset would constitute the bulk of material worth tracking (Garfield 1976). (This is justification for the limited number of journals indexed in the Web of Science.)

The question of how quality is measured and what is included in different databases is beyond the scope of this paper. We have the more modest goal of understanding scientific knowledge growth as part of a project to study why scientific collaboration (as measured by co-authorships) continues to grow. To address the question of expectations about the growth of science and why it has exceeded hypothetical expectations, we developed a systems model as to whether scientific publishing operates a complex system (Katz 2006). Showing that scientific publishing operates as a complex system (rather than simply as an aggregation of lists that continues to grow exponentially) would provide us with a set of conditions upon which we could test, understand, and further study the system for additional research. The following discussion presents the model and initial findings.

Scientific publishing has many features in common with complex systems which lead us to make this measure. The features in common include the dynamic growth of output, the emergent nature of that output, the openness of the system in accepting inputs, and the self-organization of researchers into disciplines and collaborative teams. Complex systems are dynamic and have a propensity to exhibit scaling properties along a power law form. These scaling properties appear to be a feature of scientific publication, at least in co-authorship (Wagner and Leydesdorff 2005) and in citation behavior (Katz 2000). Table 2 compares features of complex systems with features of scientific knowledge production whose statistical features have been shown by Lotka (1929), Price (1963), and others (Katz 2000) to display similarities with other complex systems.

To further test this systems hypothesis and to explain why Price’s theory of growth cannot be upheld, we constructed

---

4 Estimates have been made of the extent of scientific publication within and outside of SCIE. Bjork, Roos and Lauri estimate 2008 and Wagner and Wong’s (2011) calculation suggest that the extent of scientific publication outside SCIE may be considerably larger when the publications of developing countries are fully counted.

5 (Baranger, n.d.).
an explanatory and predictive model about why and how scientific knowledge (with publication as a proxy for knowledge) is created, diffused, and accumulated, making no other assumptions about quality.

Modeling

System dynamics enables an approach to modeling structures and simulating behaviors in cases where a structure or complex phenomena emerges. We constructed a theoretical, testable model of creation, diffusion, and accumulation and obsolescence of scientific knowledge. The system\(^6\) can be modeled as the relationship between stock and flow. Stock is conceptualized as accumulation of knowledge, which increases (or decreases) over time. The magnitude of stock is influenced by flow (inflow or outflow) over time. Relationships of such stock and flow are visualized within system dynamics as shown in Figure 4.

The stock and flow model includes the following:

- Stock, inflow, and outflow are the main variables in the structure of system.
- The stock of scientific knowledge is determined by the inflow and outflow process.
- The system is influenced by various auxiliary variables such as initial quantity, creation rate, and obsolescence rate of the stock of scientific knowledge.
- New scientific knowledge is produced by drawing upon existing scientific knowledge (which can be different or similar knowledge).
- Additional scientific knowledge is stored and adds to and replaces some scientific knowledge.
- The inflow is the combination of the recombination of existing scientific knowledge and the creation rate.
- The outflow is decided by the obsolescence rate (or death rate).
- The death rate is decided by the stock of scientific knowledge and its average life.

This scientific knowledge growth model is very similar to population growth models based on births, living persons, and deaths, which is a traditional system dynamics model. The population is the agglomeration of the number of births as inflow, the current stock of people, and the number of deaths as outflow, just as scientific knowledge consists of all the new information, accepted scientific theory along with active research, minus the scientific knowledge that no longer remains actively used or exchanged. (One key difference in the knowledge model is that some ‘obsolete’ knowledge could be revived to contribute to the system, which have been discussed in the literature as ‘sleeping beauties’ (Van Raan 2004) but might also be called ‘zombies.’) Scientific knowledge is normalized over time: some of the normalized scientific knowledge becomes generalized; and some of it disappears. In these respects, we can assume that scientific knowledge repeats this process over time with an accumulation. Based on basic assumptions, a stock-flow system model of scientific knowledge growth is constructed as seen in Figure 4.

\(^6\) A system is a set of entities (called the elements of the system) mutually related in such a way that the state of each element determines and/or is determined by the state of some other element or elements, and every element is connected by the state of some other element or elements, and every element is connected to every other by a chain of such elements. A system as a whole has a function only if it is an element in a more inclusive system, that is, only if it affects something other than itself. (Self-perpetuation does not count as a function in this sense.)
<table>
<thead>
<tr>
<th>Complex Systems</th>
<th>Scientific Publishing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynamic structure with interdependent constituents that interact in complex and</td>
<td>Adapts to new conditions (by introducing new titles); exhibits dynamism in the growth of number of titles; and non-linear in that inputs into the system are not directly related to the outputs</td>
</tr>
<tr>
<td>non-linear ways</td>
<td></td>
</tr>
<tr>
<td>‘Open’ in that new members can join, information flows across its boundaries</td>
<td>Proliferation of venues and outlets; challenge of saying what is or is not ‘scientific’ publishing; soft lines between science</td>
</tr>
<tr>
<td>which in turn are difficult to clearly identify</td>
<td></td>
</tr>
<tr>
<td>Exhibits emergent behaviors and patterns that are not caused by a single entity</td>
<td>Disciplines and sub-disciplines emerge and result in proliferation of journals based upon simple rule of seeking to claim a stake in the intellectual landscape</td>
</tr>
<tr>
<td>in the system but may arise from simple rules</td>
<td></td>
</tr>
<tr>
<td>Self-organizes in that emergent properties may change the system’s structure</td>
<td>Disciplines self-organize around theories and nomenclature, and shift the production curve upwards as new journals/articles are added</td>
</tr>
<tr>
<td>or create new structures</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.** Scientific publishing as a complex system

![Flow and stock diagram](image)

**Figure 3.** Flow and stock diagram

![Causal model of stock and flow of knowledge](image)

**Figure 4.** Causal model of stock and flow of knowledge
In the progress of science, there is some knowledge that is not superseded or nullified, no matter how many publications are published. The core knowledge is no longer cited. Thus, the box “knowledge” in Figure 4 assumes that some core human knowledge remains stable. The model shows causal relationships among variables whose interaction creates system dynamics. The system determines a behavior regarding scientific knowledge growth. The relationships between knowledge growth and its behavior can be explained in two structural features—stock and flow, and feedback loop. First, the knowledge growth can be explained as a function between net creation rate and initial knowledge as seen in Equations 1 and 2. The input of new scientific knowledge creates growth of scientific knowledge in the system which is stored, and it replaces other scientific knowledge in the Popperian sense of eliminating error or confirming earlier findings. Thus, some knowledge experiences ‘obsolescence’ in the sense that it is no longer cited even though a small percentage becomes part of normalized knowledge. The growth of knowledge can be assumed to influence the obsolescence rate (this can also be a proxy for lower quality work) with crowding increasing the rate. In this regard, the growth of knowledge can be determined by net creation rate and initial stock of knowledge. The net creation rate is determined by the gap between creation and obsolescence rates.

Knowledge

\[ \text{Knowledge} = \text{INTEGRAL (net creation rate, initial knowledge)} \]

The net creation rate is determined by the gap between creation and death rates.

\[ \text{net creation rate} = \text{creation rate} (cK) - \text{obsolescence rate} (oK) = (c - o)^a * K \]

where \( c \) is a creation rate, \( o \) is an obsolescence rate, and \( a \) is a power value of the relationship between \( c \) and \( d \), and \( K \) is amount of knowledge. Specifically, the pattern of scientific knowledge growth, pure exponential growth, or exponential growth with saturation can be determined by the pattern of net creation rate as seen in Figure 5. If creation dominates obsolescence (\( c>o \)) linearly (\( a=1 \)) or nonlinearly (\( a>1 \)), scientific knowledge will show pure exponential growth. However, if creation dominates death, and then death dominates creation like the bell-shape as seen in Figure 6, the pattern of scientific knowledge growth will show an s-shape growth (here, an exponential growth with saturation).

Scientific knowledge growth can be identified along with various feedback structures or loops around stocks and flows. Feedback structure or loops means a closed causal circle among variables. In general, there are two types of feedback structures: positive (self-reinforcement) and negative (balance). Positive feedback structures are self-reinforcing processes wherein action creates a virtuous circle. Negative feedback structure, on the other hand, is a process to stabilize or balance a system. Thus, behaviors in a system become determined by a type of feedback loop in the system. When a positive feedback dominates the whole system, the system tends to show an exponential growth. When negative feedback dominates, the system tends to show an upward or downward convex growth. An s-shape growth in a system tends to be formed when positive feedback dominates, followed by when negative feedback dominates.

In our system dynamics model, there is one positive feedback structure and two negative feedback structures as seen in Figure 5. The first positive feedback structure (R1) is the feedback loop from knowledge through creation back to knowledge.
Figure 5. Exponential growth versus saturation levels on Price’s saturation curve
The quantity of creation is determined by knowledge and creation rate positively reinforcing one another, and the size of knowledge is determined by the size of creation. In other words, the larger the knowledge the larger the creation, the larger the creation rate the larger the creation capacity, and the larger the creation capacity the larger the knowledge quantity. This feedback structure makes a self-reinforcing loop that grows scientific knowledge continuously.

There is a negative feedback loop (B1) to control the pace of knowledge growth. The negative feedback structure is the feedback loop from knowledge through normal obsolescence and obsolescence back to knowledge. Knowledge can and does become obsolescent at some point. When scientific knowledge is obsolete, articles are no longer cited. The obsolete scientific knowledge negatively influences knowledge growth. The pace of obsolescence is influenced by the normal obsolescent rate positively. In other words, the larger the normal obsolescence rate, the larger the amount of obsolescent material. However, the normal obsolescence rate is influenced by the average lifetime of scientific knowledge negatively; that is, the longer average lifetime of knowledge the fewer the normal obsolescence rate. In this respect, this negative feedback structure tends to set the pace of knowledge growth.

The size of obsolescence is also influenced by the effect of knowledge crowding on obsolescence as seen in the negative feedback loop (B2). This loop is formed by causal relationships among knowledge, normalized knowledge, effect of knowledge crowding on obsolescence, and obsolescence. Through the loop, we can see how scientific knowledge becomes obsolescent. Normalized scientific knowledge is determined by the function \( f = \text{knowledge/carrying capacity} \). In other words, scientific knowledge and carrying capacity influence the formation of normalized knowledge positively and negatively, respectively. Kuhn (1962) suggested that within the structure of scientific revelations, the stock of scientific knowledge is normalized. We accept this view; the stock is the carrying capacity of knowledge among scientists in the system plus codified knowledge. Thus, in our model, we assume that scientific knowledge repeats this process over time with an accumulation. In addition, the normalized knowledge is assumed to have a nonlinear relationship with its effect on obsolescence over time. Obsolescence is expected to increase more quickly than knowledge as knowledge grows to a level of quantity (that we do not determine in this experiment). In measurable systems, it is difficult to express the relationship between them as a constant value and it may be that the value is not constant, although we did not test for this fact.

The type of feedback loop (positive or negative feedback loop) will determine the direction of the growth of the scientific knowledge system. The dynamics will determine whether the system displays exponential growth or s-shape growth. In other words, when the positive feedback structure (R1) dominates the other two negative feedback structures (B1 and B2) in our knowledge system, it creates an exponential growth mode. The s-shape growth that Price (1963) anticipated in his scientific knowledge growth model would require negative feedback dominating at least one of the two positive feedback structures. To test the model, we constructed a simulation setting based on a stock-flow model. The simulation uses empirical data of the actual number of published articles per year between 1990 and 2010 as drawn from the Web of Science. Information of each parameter is described in the appendix.
Analysis

We ran the system dynamics model to match a systems model of exponential growth to actual scientific knowledge growth. The processes used are base-run simulation, extended simulation, and sensitivity analysis. In the base-run simulation model, we calibrated the simulation model with real data to see how well the assumptions about systemic features simulated within the model fit real data. Based on the results, we conducted several extended simulation analyses to estimate scientific knowledge growth direction in future. This section describes the simulations.

Base-run simulation

We first tested how accurately our system dynamics model calibrates current growth of published articles as shown in Figure 2. The simulation test output is based on real parameter information revealing a pattern very similar to that found by Jinha (2010). Simulation outputs of system dynamics as seen in Figure 6 and Table 2 show how simulated outputs are similar to actual data.

Figure 6 shows the similarity between the actual output and the simulation output per year. The dotted line represents our simulated output, and the solid line represents the actual data. When the two lines are compared, the simulation output is almost perfectly aligned with the actual data. Table 3 shows a comparison of actual growth of published articles with growth of the published articles through the simulation numerically one by one as seen in the table. Columns 2, 5, 8, and 11 represent the number of actual published articles per year; columns 3, 6, 9, and 12 show the outputs drawn from the simulation. The simulated number of published articles in 1991 and in 2009 reached 867,769 and 1,477,664, respectively. The actual published articles in 1991 and in 2009 are 867,807 and 1,477,383, respectively. Thus, the simulation model can be shown to have enough explanatory power in tracing the current knowledge growth to validate the model.

Extended simulation

Following validation of the base-run simulation, we simulated future growth of published articles as well as growth based on Price's (1963) scientific knowledge output model. Price anticipated that scientific articles would reach a turning point to move toward the s-shape of growth after 30 years. To test Price's hypothesis, our simulation model was only extended to 2040. As a result of our model, the number of expected published articles grew to around 3.770M as seen in Figure 7. The pattern of growth can remain at the exponential rate under the current parameters.

Sensitivity analysis

We conducted a sensitivity analysis to ensure reliability and predictive power of the model under various environmental uncertainties. In simulation models, sensitivity analysis helps to build confidence by studying the

---

7 For this aspect of testing, the annual global research output of articles investigated by Jinha (2010) draws from the pattern of scientific knowledge growth of accumulated articles between 1726 and 2009. Our simulation tested the time period between 1990 and 2010 because all real parameters between the 1726 and 1990 values could not be obtained. Thus, the base year is 1990 in our simulation model. Note that the information of each parameter is described in the appendix.
Figure 6. Actual data versus base-run simulation output between 1990 and 2009

Table 3. Comparison between actual outputs and simulation outputs

<table>
<thead>
<tr>
<th>Year</th>
<th>Actual</th>
<th>Simulated</th>
<th>Year</th>
<th>Actual</th>
<th>Simulated</th>
<th>Year</th>
<th>Actual</th>
<th>Simulated</th>
<th>Year</th>
<th>Actual</th>
<th>Simulated</th>
</tr>
</thead>
</table>

Figure 7. Extended simulation output
uncertainties that are often associated with model parameters. Even though many parameters in system dynamics models represent quantities, it is very difficult or even impossible to measure the parameters to a high degree of accuracy in the real world. Also, some parameter values change in the real world over time (Forrester 1996). In this regard, our system dynamics model was tested to see how published article growth behavior in our system dynamics model would change after altering key parameters. The sensitivity analysis also validates the model. For analyzing sensitivity in our system dynamics, two key parameters were analyzed for the sensitivity of published article growth based on average lifetime and creation rate. The sensitivity of each parameter change was analyzed by endowing each parameter with a range as follows:

\[ \text{average lifetime} = \text{RANDOM UNIFORM} (7, 8) \]

\[ \text{creation rate} = \text{RANDOM UNIFORM} (0.30, 0.35) \]

where the average lifetime is assigned between seven and eight years (as found in the real world), and the growth rate is randomly unit-distributed between the fraction per year 0.30 and 0.35. Also, where:

(1) Change current (fractional) creation rate, 0.325 up to 0.35 or down to 0.30 in the year 2009 to see what the published article growth would be in future. Growth patterns of published articles appeared as seen in Figure 8. When the creation rate was up to 0.35, the number of published articles became 5.774M; on the other hand, when the creation rate was down to 0.30, the number of published articles became 1.671M. Thus, this output shows that published article growth could be more sensitive when it was down than when it was up. We can see that the scientific knowledge output growth would be considerably daunted when creation rate was down, but it still shows a growth pattern.

(2) Change the current parameter value regarding average lifetime, 7.5 years up to 8.0 years, or down to 7.0 years in the year 2009 to see what the published article growth would be in the future. Growth patterns of published articles appeared as seen in Figure 9. When the average lifetime was up to eight years, the number of published articles became 6.737M; on the other hand, when the average lifetime was down to seven years, the number of published articles became 1.905M. This output shows that published article growth could be more sensitive when it is up than when it is down.

(3) Conduct sensitivity analysis by considering changes of both creation rate and average lifetime. The graph in Figure 9 shows the behavior of published article growth when the two parameters are changed at the same time.

The current extended simulation line is based on the parameter values creation rate = 0.35, and average lifetime = 7.5. At the extended simulation line, there exist four confidence bounds (50%, 75%, 95%, and 100%) for all the output values of published article growth shown in Figure 10. These bounds were generated when the two parameters were randomly varied around their distributions at the same time. From this sensitivity analysis, we can expect that the growth of published articles would show a purely exponential growth rate, especially when each parameter was changed to an upper value. However, when the parameters, creation rate and average lifetime were lowered below the current parameter values, the published article growth does not follow a purely exponential growth pattern.
Figure 8. Sensitivity analysis: what if creation rate was up or down

Figure 9. Sensitivity analysis: adjusting average lifetime up or down
Figure 10. Sensitivity analysis: adjusting creation rate and average lifetime change up or down
Conclusion

The rate of production of scientific publications appears to be continuing on an exponential growth curve against the prediction of Derek de Solla Price. (This article examines only publications, but it has been noted that scientific data (Borgman, Wallis, and Enyedy 2007) and e-Science (Hey and Trefethen 2005) are also growing phenomena, as well.) The growth of scientific publications has many possible causes, but the system itself appears to be operating efficiently. The networked nature of global science (Wagner and Leydesdorff 2005), the expansion of source materials and venues, the expansion of the practice of science to new places, the application of science to new problems (such as climate change), and the rise of China as a scientific power all may be contributing to the very rapid growth in output, increasing the complexity of the system.

The model constructed for this article suggests that scientific output may continue to grow exponentially. Many of the abstracting databases such as Web of Science (SCIE) are making an effort to grow their indexing services with the growth in the number of both journal titles and articles, but even so, the database used here contains only a portion of all publications (Wagner and Wong 2011). (Many of these non-source publications are in national languages and therefore remain difficult for us to access, but the materials are part of the corpus of science and can be expected to garner increasing attention over time.) Moreover, open access journals and other types of venues on the Internet continue to proliferate, shifting the face of scientific publishing to meet the needs of practitioners and users as well as the capacity of the tools to deliver information. The proliferation of sources contributes to the growth trend, and also makes it more difficult to count the outputs, but we can assume that these outputs are part of the complex system of communications of science.

Others have noted the challenge to public policy of the problems of counting scientific output for national comparisons:

It is problematic that SCI has been used and is still used as the dominant source for science indicators based on publication and citation numbers. SCI has nearly been in a monopoly situation. This monopoly is now being challenged by the new publication channels and by new sources for publication and citation counting. It is also a serious problem because a substantial amount of scientometric work and of R&D statistics has been done using a database which year for year has covered a smaller part of the scientific literature. (Larsen and Von Ins 2010, 601)

Derek de Solla Price presented the saturation hypothesis in part because exponential growth does not continue unchecked for long periods of time without one of several things happening: the physical limitations of its surroundings must change so as not to restrict growth, or the carrying capacity of the system must shift to absorb larger numbers. Both of these phenomena may be occurring. With regard to the Internet, the carrying capacity of the system has shifted to enable the absorption of new material. The rate of growth within disciplines may follow the Price’s saturation curve while the rate of growth as a whole remains exponential for the near future, and this possibility needs to be tested in future research.

Price’s intuition about the limits of the system in a state of exponential

---

The rate of growth may not be constant; it is not possible to measure the entire system.
The Price of Big Science

growth may still obtain to science but with a different interpretation. Exponential growth may have contributed to a phase shift to a new state, one characterized by abundance. Systems growing exponentially are known to reach a point of phase shift. The abundant state varies from the previous state that Price studied, where knowledge production operated within systemic features of scarcity. Paper-based output with limited availability, housed largely in the libraries of elite institutions, created conditions of a scarce system similar to biological systems scarcity. The rise of electronic sharing capabilities and broader accessibility has created conditions of abundance. In systems terms, this would contribute to systemic features of openness, sharing, reciprocity, and collaboration, but also waste and redundancy, features increasingly seen in science.

Abundant knowledge challenges governance on a number of levels. Science policy has evolved under conditions of knowledge scarcity, where nations have paid for and claimed credit for the results of research physically tied to institutions. Abundance has shifted this physicality as well as accessibility, enabling practitioners with Internet access to begin to tap the vast and historic stores of scientific publications and data as they come online. The result has been both greater collaboration and the seeding of capacity in new geographic spaces. It also increases competition for use of scientific knowledge that contributes to economic growth and competitiveness.

With the insights provided by the models presented here, policymakers can predict and control variables in ways that may optimize policy effects (i.e., knowledge creation or diffusion). These models can be applied to improve the knowledge collaboration, increase the value of resources, and result in improved use of federal research budgets, even in a time of budgetary austerity. Efficiency can be enhanced by understanding how to integrate knowledge under conditions of abundance. This means a policy shift to track research globally (to avoid redundancy) and to increase integration opportunities. It also means supporting collaborative research opportunities in which U.S. scientists and their global counterparts work to tap knowledge in many venues.

References


Borgman, C., J. Wallis, and N. Enyedy. 2007. “Little Science Confronts the Data Deluge: Habitat Ecology, Embedded

---

9 Indeed, a scarce ecosystem tends to support just a few large species, some small species, and limited vegetation—consider the northern tundra as an example. The scarce knowledge system supported only a few nations conducting science, and within them, a few elite institutions that produced the majority of products.


