# Measuring the Evolution of the Drivers of Technological Innovation in the Patent Record

Andrew Buchanan[**]
Reed College

Norman H. Packard[†]
ProtoLife Inc.

Mark A. Bedau[*,**,†]
Reed College

**Abstract** We argue that technology changes over time by an evolutionary process that is similar in important respects to biological evolution. The process is adaptive in the sense that technologies are selected because of their specific adaptive value and not at random, but this adaptive evolutionary process differs from the Darwinian process of random variation followed by natural selection. We find evidence for the adaptive evolution of technology in the US patent record, specifically, the public bibliographic information of all utility patents issued in the United States from 1976 through 2010. Patents record certain innovations in the evolution of technology. The 1976–2010 patent record is huge, containing almost four million patents. We use a patent's incoming citations to measure its impact on subsequent patented innovations. Weighting innovative impact by the dissimilarity between parent and child technologies reveals that many of the most fecund inventions are door-opening technologies that spawn innovations in widely diverse categories.

**Keywords**
Technology, innovation, patent, evolution, adaptation, citation, door-opening

*A version of this paper with color figures is available online at http://dx.doi.org/10.1162/artl_a_00022. Subscription required.*

## 1 Evolution of the Drivers of Technology

Based on an analysis of US patent records, we argue that technology undergoes an evolutionary process analogous to the process by which life evolves. Evolution is one of the hallmarks of life; many consider a system to be alive only if it is a member of an evolving population [28, 29], and some even consider such an evolving population to be itself alive [3]. So if technology changes by an evolutionary process, then technology itself (considered broadly, as the set of all patented technological artifacts) could be considered to be, in some sense, alive [1, 25], and the concept of living technology could even become meaningful [10, 11]. Since the patent record is a fairly complete and accurate history of most technological innovations, mining the patent record is a feasible and empirical way to observe how technology actually evolves. Such a study might eventually illuminate the evolution of life in all its forms.

Our goal in this article is to address three questions about the evolution of technology over the past three decades:

1. What *superstar* inventions have been the biggest drivers of the evolution of technological innovation?

* Contact author. E-mail: mab@reed.edu
** Center for Advanced Computation, Reed College, Portland, Oregon 97202, USA.
† ProtoLife Inc., San Francisco, California.

 

2. To what extent is a superstar's impact on innovation due to its unique adaptive features, as opposed to features of any randomly chosen invention?
3. Can we discern any general patterns across all patents in the unique adaptive features of superstar inventions?

We will defend the following answers to those questions: (1) The biggest drivers of technological innovation over the past four decades include three technologies: inkjet printing, polymerase chain reaction (PCR), and stents. (2) Comparison with the random selection null hypotheses shows that the evolution of technology is significantly shaped by an adaptive selection process. (3) Many of the biggest drivers of technological innovation have the general character of opening doors to rich and diverse arrays of subsequent technologies. Inkjet printing and PCR are both door-opening in this way, but stents are not.

Even if people generally agree that technology evolves, there is controversy about whether the evolution of technology is in some interesting sense "Darwinian" or "adaptive" [1, 12, 15, 22, 25, 30, 42]. Much of the same controversy arises in connection with cultural evolution in general [13, 14, 35], and especially in connection with so-called memetics [2, 17, 18]. To help navigate these controversies, we distinguish between two kinds of evolutionary processes. We use a broad conception of *adaptive* evolution, which applies to any evolutionary process that selects items from a population because of their specific adaptive value (such as the way they promote survival or reproduction). In this article we measure adaptive value as the deviation from a null hypothesis of random selection in a shadow model, as explained below in Section 3.1. *Natural selection* by definition occurs whenever a population of entities has heritable and variable traits that affect the probability of those entities surviving and reproducing (see, e.g., [18, 27]). This is the process that tends to proliferate those heritable and variable traits that make organisms more likely to survive and reproduce. In this article, we say that a population is undergoing *Darwinian* evolution if and only if natural selection is a significant factor in explaining how certain traits in the population change over time. A population undergoing Darwinian evolution can simultaneously be affected by non-Darwinian processes (e.g., random genetic drift, genetic synergies, and epigenetic factors [23, 38]), provided that natural selection is *among* its significant explanatory factors. By definition, a Darwinian evolutionary process is an adaptive evolutionary process, but not necessarily vice versa, for natural selection is not the only process that can select items for their adaptive value.

Below we present evidence that the evolution of the drivers of innovation is adaptive, and not a result of random selection. This adaptive evolutionary process skews the distribution of traits in the evolving population of technology drivers. Natural selection also skews the evolving distribution of traits in certain reproducing populations. But the adaptive evolution of technology drivers is *not* necessarily due to natural selection, at least not if the Darwinian fitness of a particular invention is measured by its market share, which many find natural [42]. Consider the hypothetical example of a patented invention that is not widely adopted in the market but is cited by many subsequent patented inventions. This invention has negligible Darwinian fitness, but is still a significant driver of technology. Because an unfit Darwinian patent could be a leading driver of the evolution of technology, the evolution of the drivers of innovation is not merely the process of natural selection.

Given comprehensive data on the market share of different technologies over time, one could compute the evolutionary activity statistics that Bedau and Packard have applied to many biological populations and artificial life models [7–9, 32, 33]. The empirical evolutionary activity statistics would help resolve whether the evolution of technology is Darwinian and involves natural selection. Here, rather than analyzing market data, we use citation counts to measure an invention's impact on subsequent innovation. The key measure for us is whether a patent's impact is higher than what you would expect if citations were distributed randomly. Our impact statistics can be viewed as a biologically inspired development of previous work that uses citation statistics to characterize the knowledge economy [24], and previous work on statistics and models for networks of citations among published scientific articles [34, 26] and the patent record [40].

Although technological and biological evolution are different in many respects, the two processes are still analogous. Some scientists place both processes under one general abstract model of evolution (e.g., [41, 42]) or of selection [19, 21]. Just as one can view a technology's market share as its Darwinian fitness, one can view the production of technological artifacts as analogous to reproduction of individuals in a biological population. The process by which a new, better technology spreads and becomes widely adopted by people is analogous to the process by which a dominant species invades a niche. The spread and adoption of innovations has been modeled as a diffusion process [36], which can be affected by people's differential ability to retain information [39].

Although this article stresses a certain similarity between the evolution of technology and that of biology, the two processes have some important differences. One is the way in which innovations originate. Conscious intentions of different human agents play a significant role only in technological innovation (e.g., [1, 16, 20, 42], but see also [25]), and various models of the generation of innovations have been proposed (e.g., [36, 42]). Brian Arthur [1] has stressed a second difference: new technologies can be created by freely combining previous technologies in a way that is usually impossible in biological evolution. A third notable difference between patents as an evolving system and biological evolution follows from the fact that citations may be mandated by patent examiners as part of the patent approval process, and inconsistency in patent examiners may attenuate the reliability of citations as evidence of intellectual lineage. In general, lineages among patents appear more arbitrary and artificial than biological lineages. Fourth, the evolution of technology involves no clearly separable generations. Any patent may be cited at any time in the future. One patent may be a grandchild in the sense that it refers to a patent that refers to the original parent patent, but it may also refer to the original parent patent, so it would simultaneously be a child and a grandchild of one and the same patent. So, patent lineages are unusually tangled. Fifth, since there is usually no direct copying of text from parent to child, the notion of variation based on mutation and recombination cannot be the same in biology and technology. In the population of patents, variation is ensured by the fact that patents are granted only to sufficiently novel innovations, so each patent must include important new traits along with those it inherited. In addition, when an inherited trait is put into the new context of the child patent, the meaning or function of the inherited trait can be different in the second context; that is, the meaning or function of a technological trait is context-sensitive. Sixth, selection and death do not have the same meaning as in biological evolution; selection among patents is indicated by differential citation of some patents over others (which we demonstrate below), and death is simply never being cited by future patents. Since one can never be sure that a patent will never be cited in the future, hitherto uncited patents can be confidently classified only as dormant. This contrasts with the certain extinction that afflicts biological species. Finally, as a consequence of the novelty required for each patent, every new innovation is necessarily significantly different from the earlier patents that it cites, so patented inventions do not reproduce in kind in the way that biological populations do.

All of these disanalogies notwithstanding, we maintain that the changing patent population is still an evolutionary process in which heritable traits pass by means of citations to new members of the patent population. We will show that this evolutionary process is adaptive, in the sense that citation of patents is nonrandom and depends on a patent's specific traits. We use a flexible suite of impact statistics to measure the degree to which selection is nonrandom, and we normalize those statistics with a variety of heuristics. One robust conclusion revealed by those statistics is that PCR, inkjet printing, and stents in recent decades have been superstars at driving technological innovation. We conclude by showing how the statistics indicate that door-opening inventions are significant drivers of technological innovation.

## 2 The Patent Record

The patent data we mine in this experiment consist of records of all US utility patents issued over thirty-four years from 1976 through 2010. This patent record is analogous to the fossil record for those who study the evolution of life. Figure 1 shows that over the past thirty years the rate at which patents have been issued has more than doubled. In this study we focus on a few key pieces of information in the patent record: patent number, title, issue date, International Patent Classification (IPC), and references. The patent number is a unique identifier for each patent.

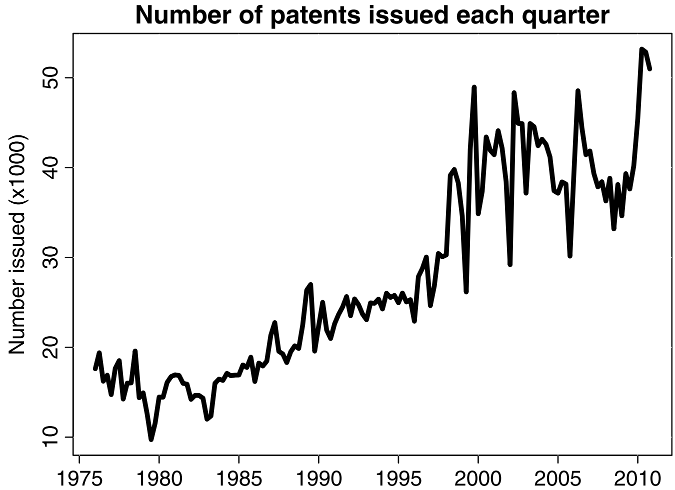


Figure 1. Number of patents issued each quarter, from 1976 through 2010.

Each US patent is assigned a handful of IPCs by the inventor and patent examiners at the United States Patent and Trademark Office (USPTO). The IPCs are designed to categorize the invention. In this article we use IPCs to measure the degree of similarity and dissimilarity between inventions, and to control for differences in citation practices across categories of inventions.

The front page of a patent record is required by the USPTO to cite all of the previous inventions on which the newly patented invention depends. These citations establish the invention's *prior art* and are compiled by both patent examiners at the USPTO and the inventor. Figure 2 shows a fourfold rise in the average number of citations made by each patent over the past thirty years. Citations play a pivotal role in our evolutionary analysis of the patent data. We develop a precise formalism for key statistics about citations, and visualize the evolution of technology by plotting the time course of the most heavily cited inventions.

In this study of technological evolution, for simplicity we often identify a patented invention with its patent record. A patent's traits are the features of the invention disclosed in its patent record. Since each new patent cites the important previous patents on which it depends, a child patent shares

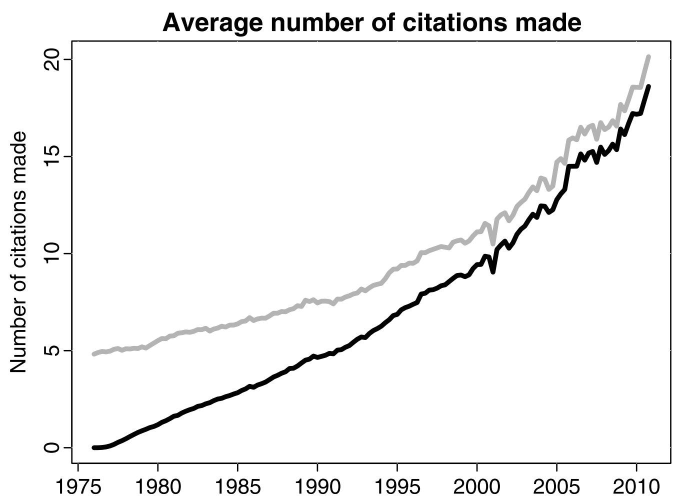


Figure 2. Average number of citations made by patents issued each quarter. Upper curve includes all citations made; lower curve includes only citations made to patents within our data set.

important traits with each of the parent patents it cites. In biological evolution, traits are inherited when genes are passed from parent to child. In the evolution of the drivers of technology, citations provide indirect observable empirical evidence of the common intellectual content shared by parent and child innovations, and thus disclose intellectual lineages. Further work can reveal which key traits are driving the evolution of technology at any given time [15].

## 3 Measuring Innovative Impact

We define the *impact* on subsequent technological innovation of a patent as the cumulative number of times other patents cite it. For patent $p$, $c^t(p)$ is defined as the set of patents issued at time $t$ that cite $p$, and $C_p^t$ as the cumulative citations to patent $p$ up to $t$:

$$C_p^t = \sum_{t'=0}^{t'=t} \sum_{p' \in c^{t'}(p)} f^t(p, p'), \tag{1}$$

where $f^t(p, p')$ is a counting function, constructed to count contributions of citations to the cumulative sum. The simplest version of a counting function is $f^t(p, p') \equiv 1$, in which case each citation in $c^t(p)$ is counted with equal weight. $C_p^t$ with this counting function is shown in Figure 3. We show below how to craft the counting function $f^t(p, p')$ to emphasize or deemphasize different aspects of the population. Impact statistics formally belong to the family of *evolutionary activity* statistics [6–9, 32, 33].

Figure 3 overlays the patent number and title for the twenty most heavily cited patents in our data set. In this and all similar plots of impact statistics in the patent record, citation impact waves appear as follows (colors available in electronic editions): Inkjet printing patents are thick dark gray (blue) lines, polymerase chain reaction (PCR) patents are thick medium gray (red) lines, and stents patents are thick light gray (green) lines. All other patents are shown as thin lines in various shades of gray. We focus on PCR, inkjet printing, and stents because those three areas of technology capture all of the 10 most heavily cited patents in Figure 3. Below we ask why the impact of those specific technologies is so high.

The average behavior of $C_p^t$, obtained by averaging over all patents issued at each new time $t$, is illustrated in Figure 4 (the time resolution is quarterly). Notice that the curves are roughly straight lines, indicating that

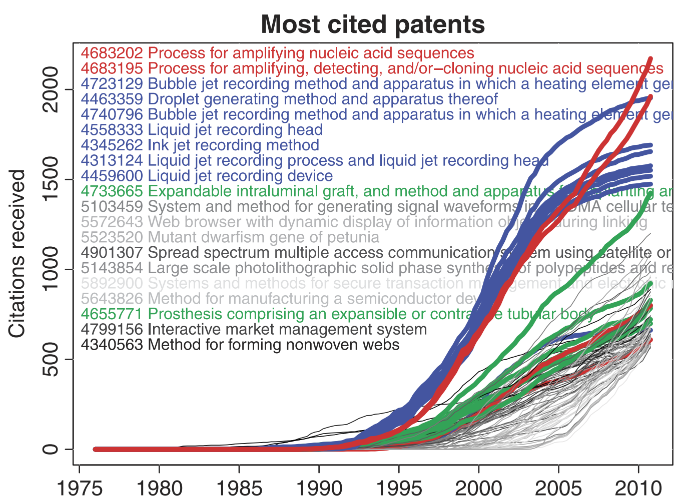


Figure 3. Impact measured by cumulative number of citations a patent receives, as a function of time. Each curve shows citations accumulated by a single patent. Only the top 100 patents are shown. Patent numbers and titles of the 20 most-cited patents appear in the same color as the corresponding curve. Note that PCR, inkjet printing, and stents stand out as superstars.

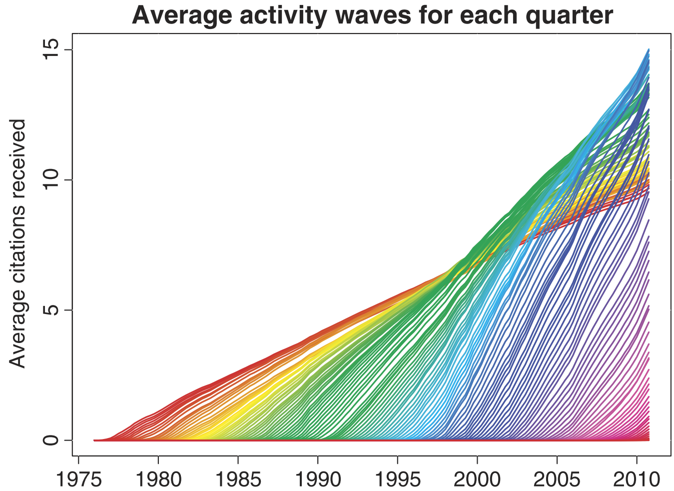


Figure 4. Average number of citations received each quarter, for each patent cohort. Each curve is the cumulative sum of the citations received by a patent in a quarter, averaged over all patents in each cohort. A cohort is the set of all patents issued in a given quarter. Colors reflect the sequence of patent cohorts.

patents continue to receive citations at roughly the same rate over their life in the data set. Note also that the slopes of the lines increase through the first two decades in our data and then level off.

## 3.1 Shadow Patents

Following a long history of constructing and using shadow models of biological evolution [4, 8, 9, 37], we construct a *shadow patent* model, in order to determine if our citation statistics could be created by citing patents at random. A shadow patent model is constructed to mirror certain aspects of the real patents, except that citations are bestowed randomly, with equal probability for all previously issued patents.

Because they "shadow" real patents, shadow patents have many of the same statistics. If a real patent is issued, then so is a shadow patent, and if a real patent makes a citation, then so does a shadow patent, so Figures 1 and 2 are necessarily identical for real and shadow patents. However, citation statistics for real and shadow patents can differ, so Figure 3 for real and shadow patents can differ. When shadow patents select *which* earlier patents to cite, they do so *randomly* and with equal probability from the pool of available earlier patents.

To test the hypothesis that heavily cited real patents are heavily cited just by chance (given the number of patents being issued and the number of citations being made), we simulate shadow patents and observe their typical maximal citation levels. If the heavily cited real patents have significantly more citations that the most-cited shadow patent, then heavily cited real patents are not just statistical fluctuations. They are driving technological innovation because of their specific adaptive features.

Figure 5 shows the cumulative citations of the 100 most heavily cited shadow patents, analogous to Figure 3. Comparison of the *y* axes in Figures 3 and 5 shows that the most heavily cited real patents have nearly two orders of magnitude higher citation levels than the most heavily cited shadow patents. This shows the nonrandom nature of the process by which especially fecund inventions seed a great many later innovations. The variance in citation levels among real patents is not mere noise. Rather, there is something special about highly cited patents that makes them so influential. This itself is direct evidence that the evolution of technology is adaptive, as we have defined it.

## 3.2 Superstar Patents

Further insight into high-impact inventions may be gained by examining the most highly cited patents—which we term patent *superstars*. Creating narratives about superstar patents can contribute to our intuition about what drives the evolutionary process. Further examination of Figure 3 reveals that superstar patents typically involve one of the following three innovations: PCR, inkjet printing, and stents. Examining these inventions explains why they each made such a large impact on the future evolution of technology.

### 3.2.1 PCR

Polymerase chain reaction is one of the cornerstones of contemporary biotechnology. Patented in 1987 by Kary Mullis of Cetus Corporation (one of the first biotech firms), PCR makes it possible to rapidly make millions of copies of a given DNA molecule. Among other things, this makes it possible to identify conclusively the sequence in a microscopic sample of DNA. This method has been extensively modified to achieve many different kinds of genetic manipulations. It is now a fundamental tool in a wide range of biotech applications. In 1993, Mullis received the Nobel Prize in chemistry for his work on PCR.

### 3.2.2 Inkjet Printing

The Japanese company Canon holds a series of patents on inkjet printing that have been very heavily cited. Although originally developed for putting ink on paper, the fundamental innovation behind inkjet printing actually involves something much more general: the ability to position extremely small bits of matter (ink) with extreme precision. Beyond traditional dyes used for printing on paper, the materials now used also include skin cells (so skin grafts can be printed), DNA or RNA primers (which are printed on microarray chips), and even metals. Depositing successive layers of materials means that we can print countless different kinds of three-dimensional objects. One now reads about inkjet printing technology being used to print batteries, clocks, and flexible video screens, among other things.

### 3.2.3 Stents

Stents are man-made tubes that are used to hold open conduits in the body, such as coronary arteries partially occluded with plaque. In 1986, Julio Palmaz patented a stent that could be expanded within a blood vessel by an inserted angioplasty balloon. This procedure allows some blocked coronary arteries to be repaired without open-heart surgery, making medical treatment much simpler and safer. Citations

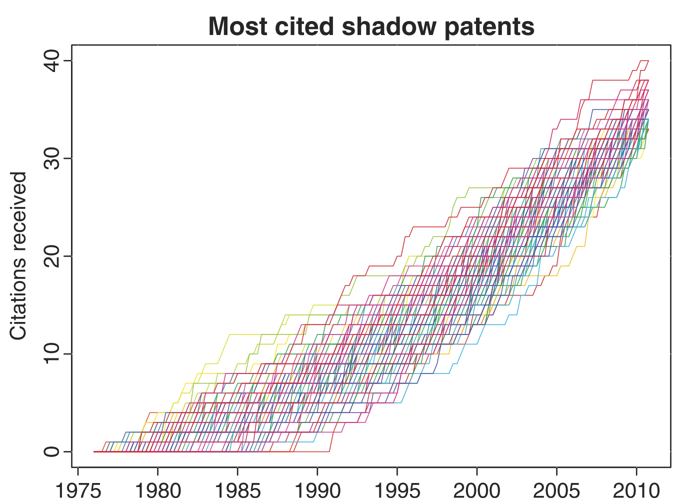


Figure 5. Impact measured by cumulative number of citations of the 100 most heavily cited shadow patents. Colors reflect individual shadow patents. Compare with the most heavily cited real patents shown in Figure 3.

to this patent indicate that it led to a range of new kinds of minimally invasive blood vessel therapies. Stents have been in the news recently because of patent litigation between Boston Scientific and Johnson and Johnson, and because of controversy about the merits of drug-coated stents.

Although raw citation counts from 1976 to 2010 identify PCR, inkjet printing, and stents as superstar inventions, there are other superstar inventions during the same period, of course. Using similar methods to those applied here, Chalmers et al. [15] provide evidence that semiconductors, e-commerce, and wireless communication, for example, are also among the significant drivers of innovation during the last few decades. It is easy to adapt these methods to highlight the drivers of innovation in specific commercial sectors, geographical areas, countries, and the like. In addition, we will see below how various ways of normalizing raw citation counts can reveal novel perspectives on the impact of different superstar patents. But the dominance of PCR, inkjet printing, and stents remains consistent throughout.

### 3.3 Eliminating Data Biases and Artifacts

The definition of impact on subsequent innovation in terms of the raw cumulative citation counts $C_p^t$ as described above may suffer from artifacts in the data that are not related to the kind of innovative evolution that our statistics aim to capture. This leads to modifications in the definition of activity, obtained by modifying $C_p^t$ to counter these effects through a process of normalization. The canonical way in which $C_p^t$ will be modified is through the definition of the counting function $f^t(p, p')$. We will see how modified counting functions will enable biases and artifacts to be compensated for explicitly. In some cases, these modifications may contain a parameter that must be chosen for a certain level of compensation; for this reason, these modified counting functions may be regarded as heuristic, rather than fundamental.

A simple example of such an artifact is evident from Figure 2, in which the number of citations grows with time. Assuming that this increase is due to cultural changes and is not an inherent feature of the system, this results in later patents having undue weight because they make more citations.

A normalization to adjust for this effect uses the counting function

$$f_{\text{rate}}^t = \frac{M^{t'}}{M^t}, \tag{2}$$

where $M^t$ is the mean number of citations made by patents at time $t$, and $t'$ is the (arbitrary) baseline time point in the data set, 2010 in this case. The effect of this normalization is to value all citations in terms of the baseline citation rate, similar to adjusting historical prices for inflation. Because patents at the beginning of the data set make roughly one-quarter as many citations as those at the end, their citations are given roughly four times as much weight. Then, the adjusted cumulative citation sum, $C_{\text{rate }p}^t$, is computed from Equation 1 using $f^t(p, p') \equiv f_{\text{rate}}^t$.

The dynamics of $C_{\text{rate }p}^t$ are illustrated in Figure 6. Note that this normalization significantly boosts the citation counts for earlier patents, as expected. But the same 10 patents involving PCR, inkjet printing, and stents still occupy the top 10 positions in the graph. Normalizing by prior expected probability of being cited changes the relative significance of patents, but technological evolution remains most strongly driven by innovation in PCR, inkjet printing, and stents.

The average citation rates of different patent classifications vary by orders of magnitude (data not shown). These skewed IPC citation distributions could create further artifacts in our cumulative citation statistics. We can test that hypothesis by introducing a new counting function, $f_{\text{IPC}}$, to normalize by the mean number of citations made by patents in a given category.

The IPC of a patent has five levels, $I(p) = (c_1,\ldots, c_5)$, where each $c_i$ may be thought of as an integer labeling different categories. So, to define the new counting function, we first define the categories of interest to be all possible values of the first two category coordinates, $\mathbf{c} = (c_1, c_2)$. (As it happens,

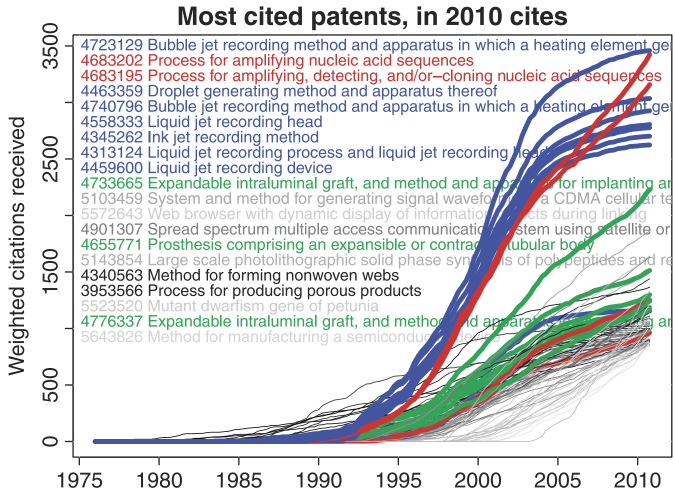


Figure 6. Normalization by relative outgoing citation rate due to changes in the number of citations being made over time. Impact is valued in terms of most-recent citation rates. Notice that PCR, inkjet printing, and stents remain superstars, and that the relative ranking of stents remains roughly the same, while the relative ranking of inkjet printing increases relative to Figure 3.

there are 129 such second-level IPC categories.) The total number of citations made by patents in the category at time $t$ is

$$R_{\mathbf{c}}^{t} = \sum_{p' \in p} r(p')\delta(c_1 - I(p')_1)\delta(c_2 - I(p')_2),$$

and the mean number of citations made is

$$M_{\mathbf{c}}^{t} = R_{\mathbf{c}}^{t}/N_{\mathbf{c}}^{t},$$

where $\delta(x) = 1$ if $x = 0$ and 0 otherwise, $r(p')$ is the number of citations made by $p'$, and $N_{\mathbf{c}}^{t}$ is the total number of patents in category $\mathbf{c}$ at $t$. So we can define $f_{\mathrm{IPC}}$ to be a function that depends only on the citing patent:

$$f_{\mathrm{IPC}}^{t}(p') = \sum_{\mathbf{c}} \frac{M_{\mathbf{c}'}^{t'}}{M_{\mathbf{c}}^{t}} \delta(c_1 - I(p')_1)\delta(c_2 - I(p')_2). \tag{3}$$

For example, a patent in category A01 issued in 1976 has its outgoing citations doubled in weight, because those patents made half as many citations on average as the baseline rate (chosen arbitrarily to be the rate for B68 patents from 2010). At each given moment this equalizes the total evolutionary activity from each patent in each category.

Figure 7 shows a plot of $C_{\mathrm{IPC}p}^{t}$, defined by Equation 1, with $f^{t}(p, p') \equiv f_{\mathrm{IPC}}^{t}(p')$. The main difference from Figure 6 is that the relative rank of PCR patents is raised. Nevertheless, the same three narratives (PCR, inkjet printing, and stents) still play a dominant role in driving technological innovations.

Another important effect present in the patent record is that some patents are cited by subsequent patents with the same assignee; we refer to this as *self-citation*. Some citation counts are significantly affected by self-citations. If a company makes an innovation, it is motivated to patent further developments of

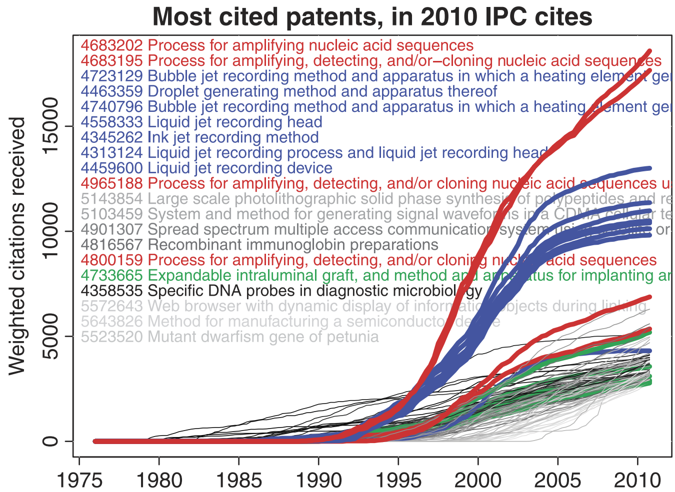


Figure 7. Normalization by mean outgoing citation rate for patents in each second-level IPC category. Notice that the relative rankings of PCR, inkjet printing, and stents change, but all remain superstars.

that innovation, thus inflating citations for patents from the same assignee. A simple normalization to filter out this effect uses a counting function that discounts self-citations, as follows:

$$f_{\text{self}}(p, p') = \begin{cases} \alpha & \text{if } p \text{ and } p' \text{ have the same assignee,} \\ 1 & \text{otherwise,} \end{cases}$$

with $\alpha < 1$. Then, the adjusted cumulative citation sum, $C^t_{\text{self}(p,\,p')}$, is computed from Equation 1 using $f^t(p, p') \equiv f^t_{\text{IPC}}(p') f_{\text{self}}(p, p')$, where we include normalization with respect to changing mean IPC citation rates, as described above for $f^t_{\text{IPC}}$.

Figure 8 shows a plot of $C^t_{\text{self}}(p, p')$ for $\alpha = 0.33$ (other values of $\alpha$ produce similar results). This normalization significantly reshuffles the relative impact of the top patents. One effect is the dramatic drop in inkjet printing patents (blue). Those patents cover inventions developed at Canon, and numerous

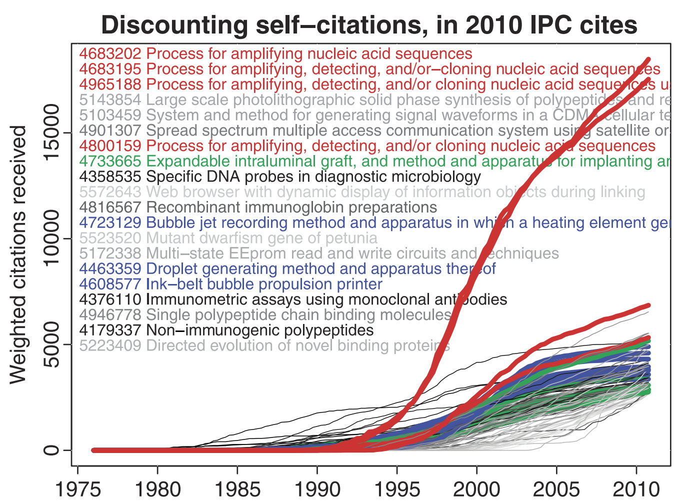


Figure 8. Discounting for self-citations and inflation (in IPC category 2010 cites). Notice that the ranking of inkjet printing dramatically falls, but inkjet printing, PCR, and stents all remain superstars.

subsequent Canon patents cite their earlier inventions as prior art. However, relatively few other groups cite Canon's inkjet printing patents. By contrast, the PCR and stent patents are largely unaffected in both relative and absolute terms.

We may combine any or all these normalizations, aiming to obtain the cleanest possible picture of which technologies most strongly drive innovation in the evolution of technology. The various combinations of normalizations that we have observed displayed a consistent pattern: PCR, inkjet printing, and stents remain among the most fecund technologies. While removing artifacts from cumulative citation counts does noticeably alter the relative ranking of the three stories, it is striking that the same drivers predominate.

## 4 Door-Opening Innovations

One might hypothesize that a crucial aspect of biological evolution is the ability of biological innovations to open doors to entire new domains of biological innovation [5], perhaps through the creation of new modes of interaction and new ecological niches [31]. Door-opening innovations contrast with inventions that achieve merely incremental progress in one limited domain. This section presents empirical evidence that door-opening innovations play an important role in the evolution of technology.

To address this issue, we simply adjust our citation statistics in the right way, once again applying the method for eliminating biases and artifacts in the patent record: We define a new counting function—a counting function that emphasizes or accentuates the property being investigated—and then we look for patterns when we plot the impact statistics. This use of a counting function is heuristic, in the sense that the function has no fundamental formulation but rather can be captured in a range of different ways, depending on the hypothesis being tested and the test being employed.

We use IPC categories to quantify the degree to which a patent is door-opening, as the breadth of IPC categories of the patents that cite it. The intuition is that if a patent is cited by patents from very similar IPC categories, then it has a narrow impact, whereas patents that are cited from many distant IPC categories have a much broader impact and are opening doors to many new areas of innovation. This intuition may be quantified by weighting the citation count more heavily for more distant IPC categories.

Specifically, if $I(p)$ is the IPC vector $(c_1, \ldots, c_5)$, with $c_1$ being the coarsest grain IPC resolution, and $c_5$ being the finest grain resolution, we define the IPC distance between two patents as

$$d_{\mathrm{IPC}} = 5 - n_{\mathrm{IPC}},$$

where $n_{\mathrm{IPC}}$ is the maximum integer such that $I(p)_i = I(p')_i$ for all $i \leq n_{\mathrm{IPC}}$. Then we may create a counting function that weights by this distance, exponentiating it to emphasize the effect:

$$f^{t'}_{\mathrm{IPCd}}(p, p') = 2^{d_{\mathrm{IPC}}}. \tag{4}$$

Now, we can compute $C^t_{\mathrm{IPCd}\,p}$ from Equations 1 and 3, using $f^t(p, p') \equiv f^t_{\mathrm{IPC}}(p') f^t_{\mathrm{IPCd}}(p, p')$.

The plot of $C^t_{\mathrm{IPCd}\,p}$ in Figure 9 shows that PCR and inkjet printing remain significant innovations when citations are weighted by their IPC distance. This provides evidence that PCR and inkjet printing have been door-opening innovations. For if those inventions spawned merely incremental progress in the same vein, then weighting citations by IPC distance would deflate their relative citation levels. But instead, those patents remain superstars in Figure 9. So, a relatively high number of their citations come from widely different kinds of technologies. In other words, PCR and inkjet printing are highly door-opening. Furthermore, citation counts for inkjet printing patents are especially boosted when weighted by IPC distance. This implies that inkjet printing has been especially effective at opening doors to a far-flung variety of further innovations. This conclusion is consistent with the wide range of different kinds of three-dimensional objects that can be fabricated today by sophisticated developments of inkjet printing.

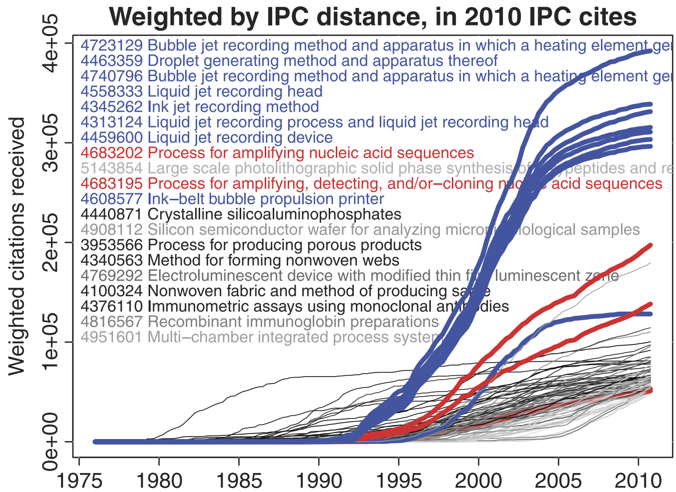


Figure 9. Weighting citation counts by the exponential of IPC distance (in IPC category 2010 cites). Weighting citations from patents in distant IPC categories rewards door-opening innovations and penalizes innovations that spur only similar types of innovation.

Emphasizing door-opening inventions impacts stents to a far greater degree than the other superstar patents, dropping them out of the top hundred patents shown. This implies that stents have not been in the same league at opening doors as inkjet printing and PCR. Intuitively this makes sense. Stents have had application only in medical procedures involving vascular interventions, primarily endovascular surgery, and they have not spawned a wide variety of kinds of inventions.

## 5 Conclusions

Our results show that technology undergoes an innovative evolutionary process. The set of issued patents can be viewed as an evolving population of entities from which new patents emerge. We use citation statistics to measure an invention's impact on subsequent innovations. The evolution of impact is different from the Darwinian process of preferential adoption of "the fittest" technologies by technology users. The innovations reflected in the patent record may well give rise to episodes of Darwinian evolution, as better and worse technologies gain and lose market share, but we ignore that Darwinian process here. Instead, we track which inventions most heavily impact the evolution of future inventions. This non-Darwinian process presumably also often occurs in biological evolution, so our analysis might eventually help elucidate evolutionary processes more generally.

We interpret cumulative citation count as impact on subsequent innovation. The dramatic citation counts for the most-cited patents show that high impact cannot be explained merely as a statistical fluctuation. The comparison with the shadow model's random-selection null hypothesis is compelling evidence for the adaptive value of the inventions with the highest impact.

Cumulative citation rates across the entire population of patents support the population-level argument for adaptive evolution of technology. This conclusion is reinforced by examining individual *superstar* patents that are cited exceptionally often. It turns out that the narratives for the superstar patents corroborate the intuitive hypothesis that the innovations observed in the patent record are driven by the specific adaptive advantages offered by the superstar inventions.

The cumulative citation count on which this conclusion is based can be adjusted to account for biases inherent in the data. We have discussed various such adjustments, and we find that the evidence for adaptive innovative impact is consistently and strongly present over all versions of adjustments we have examined. Deciding which adjustments to make is delicate, and the decision can have a substantial effect on the particular patents that emerge as superstars, and on the narratives that accompany them. Some of the difficulties are inherent in the data. For example, since patents can be cited only by subsequent patents, the latest patents in our data set have not yet received any citations.

Further heuristic adjustments to our cumulative citation count statistics may be made to emphasize or uncover certain structure in the data. We have used one such adjustment, exponential boost of citations that cross IPC boundaries, to discover which patents appear to be issued for door-opening technologies, that is, those that enable a broad range of further kinds of innovations in areas different from the original area the patent was issued in. Examining this adjustment corroborates the hypothesis that many patent superstars are for door-opening technologies. An extensive body of empirical evidence for this and similar conclusions can be gathered and evaluated by wider applications of the methods developed and illustrated here.

## Acknowledgments
Thanks to Devin Chalmers, Cooper Francis, and Noah Pepper for discussions about how to quantify the evolution of technology. For helpful comments, thanks to Emily Parke, Federico Vaggi, and the anonymous reviewers for Artificial Life XII, where we presented a preliminary version of our results and conclusions.

## References

1. Arthur, W. B. (2009). *The nature of technology: What it is and how it evolves*. New York: Simon & Schuster, Free Press.
2. Aunger, R. (Ed.) (2000). *Darwinizing culture*. Oxford, UK: Oxford University Press.
3. Bedau, M. A. (1996). The nature of life. In M. Boden (Ed.), *The philosophy of artificial life* (pp. 332–357). Oxford, UK: Oxford University Press.
4. Bedau, M. A. (2003). Objectifying values in science: A case study. In P. Machamer & G. Wolters (Eds.), *Science, values, and objectivity* (pp. 190–219). Pittsburgh, PA: University of Pittsburgh Press.
5. Bedau, M. A. (2009). The evolution of complexity. In A. Barberousse, M. Morange, & T. Pradeu (Eds.), *Mapping the future of biology: Evolving concepts and theories* (pp. 111–130). Berlin: Springer.
6. Bedau, M. A., & Brown, C. T. (1999). Visualizing evolutionary activity of genotypes. *Artificial Life*, *5*, 17–35.
7. Bedau, M. A., & Packard, N. H. (1992). Measurement of evolutionary activity, teleology, and life. In C. Langton, C. Taylor, D. Farmer, & S. Rasmussen (Eds.), *Artificial Life II* (pp. 431–461). Redwood City, CA: Addison-Wesley.
8. Bedau, M. A., Snyder, E., Brown, C. T., & Packard, N. H. (1997). A comparison of evolutionary activity in artificial evolving systems and in the biosphere. In P. Husbands & I. Harvey (Eds.), *Proceedings of the Fourth European Conference on Artificial Life* (pp. 125–134). Cambridge, MA: MIT Press.
9. Bedau, M. A., Snyder, E., & Packard, N. H. (1998). A classification of long-term evolutionary dynamics. In C. Adami, R. Belew, H. Kitano, & C. Taylor (Eds.), *Artificial Life VI* (pp. 228–237). Cambridge, MA: MIT Press.
10. Bedau, M., Hansen, P. G., Parke, E., & Rasmussen, S. (Eds.) (2010). *Living technology: Five questions*. San Diego, CA: Automatic Press/VIP.
11. Bedau, M. A., McCaskill, J. S., Packard, N. H., & Rasmussen, S. (2010). Living technology: Exploiting life's principles in technology. *Artificial Life*, *16*, 89–97.
12. Benzon, W. (1996). Culture as an evolutionary arena. *Journal of Social & Evolutionary Systems*, *19*, 321–365.
13. Boyd, R., & Richerson, P. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.
14. Cavalli-Sforza, L., & Feldman, M. W. (1981). *Cultural transmission and evolution: A quantitative approach*. Princeton, NJ: Princeton University Press.
15. Chalmers, D., Francis, C. C., Pepper, N., & Bedau, M. A. (2010). High-content words in patent records reflect key innovations in the evolution of technology. In *Proceedings of Artificial Life XII* (pp. 838–845). Cambridge, MA: MIT Press.
16. Christensen, C. M. (2000). *The innovator's dilemma: The revolutionary book that will change the way you do business*. New York: HarpersBusiness.

17. Dawkins, R. (1989). *The selfish gene* (2nd ed.). Oxford, UK: Oxford University Press.
18. Dennett, D. C. (2006). *Breaking the spell: Religion as a natural phenomenon*. New York: Viking.
19. Godfrey-Smith, P. (2009). *Darwinian populations and natural selection*. Oxford, UK: Oxford University Press.
20. Hughes, T. P. (2004). *Human-built world: How to think about technology and culture*. Chicago: University of Chicago Press.
21. Hull, D. L., Langman, R. E., & Glenn, S. S. (2001). A general account of selection: Biology, immunology, and behavior. *Behavioral and Brain Sciences*, *24*, 511–573.
22. Jablonka, E. (2002). Between development and evolution: How to model cultural change. In M. Wheeler, J. Ziman, & M. A. Boden (Eds.), *The evolution of cultural entities* (pp. 27–41). Oxford, UK: Oxford University Press.
23. Jablonka, E., & Lamb, M. J. (2005). *Evolution in four dimensions: Genetic, epigenetic, behavioral, and symbolic variation in the history of life*. Cambridge, MA: MIT Press.
24. Jaffe, A. B., & Trajtenberg, M. (2002). *Patents, citations, and innovations: A window on the knowledge economy*. Cambridge, MA: MIT Press.
25. Kelly, K. (2010). *What culture wants*. New York: Viking.
26. Lehmann, S., Jackson, A. D., & Lautrup, B. (2005). Life, death and preferential attachment. *Europhysics Letters*, *69*, 298–303.
27. Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics*, *1*, 1–18.
28. Maynard Smith, J. (1975). *The theory of evolution* (3rd ed.). New York: Penguin.
29. Mayr, E. (1982). *The growth of biological thought*. Cambridge, MA: Harvard University Press.
30. Mokyr, J. (1990). *The lever of riches: Technological creativity and economic progress*. Oxford, UK: Oxford University Press.
31. Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (2003). *Niche construction: The neglected process in evolution*. Princeton, NJ: Princeton University Press.
32. Raven, M. J., & Bedau, M. A. (2003). General framework for evolutionary activity. In W. Banzhaf, T. Christaller, P. Dittrich, J. Kim, & J. Ziegler (Eds.), *Advances in artificial life, 7th European Conference ECAL 2003* (pp. 676–685). Berlin: Springer.
33. Rechtsteiner, A., & Bedau, M. A. (1999). A generic model for quantitative comparison of genotypic evolutionary activity. In D. Floreano, J.-D. Nicoud, & F. Mondada (Eds.), *Advances in artificial life* (pp. 109–118). Berlin: Springer.
34. Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B*, *4*, 131–134.
35. Richerson, P. J., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago: University of Chicago Press.
36. Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.
37. Skusa, A., & Bedau, M. A. (2002). Towards a comparison of evolutionary creativity in biological and cultural evolution. In R. Standish, M. A. Bedau, & H. A. Abbass (Eds.), *Artificial life VIII* (pp. 233–242). Cambridge, MA: MIT Press.
38. Sperber, D. (1996). *Explaining culture: A naturalistic approach*. New York: Blackwell.
39. Strimling, P., Enquist, M., & Eriksson, I. (2009). Repeated learning makes cultural evolution unique. *Proceedings of the National Academies of Sciences of the U.S.A.*, *106*, 13870–13874.
40. Valverde, S., Sole, R. V., Bedau, M. A., & Packard, N. (2007). Topology and evolution of technology innovation networks. *Physical Review E*, *76*, 056118.1–056118.7.
41. Wheeler, M., Ziman, J., & Boden, M. A. (Eds.) (2002). *The evolution of cultural entities*. Oxford, UK: Oxford University Press.
42. Ziman, J. M. (Ed.) (2000). *Technological innovation as an evolutionary process*. Cambridge, UK: Cambridge University Press.