Age and Great Invention*

Benjamin F. Jones†

March 2007

Abstract

Great achievements in knowledge are produced by older innovators today than they were a century ago. Using data on Nobel Prize winners and great inventors, I find that the mean age at which noted innovations are produced has increased by 6 years over the 20th Century. I estimate shifts in life-cycle productivity and show that innovators have become especially unproductive at younger ages. Meanwhile, the later start to the career is not compensated for by increasing productivity beyond early middle age. I further show that the early life-cycle dynamics are closely related to variation in the age at Ph.D. and discuss a theory where accumulations of knowledge across generations lead innovators to seek more education over time. More generally, the results show that individual innnovators are productive over a narrowing span of their life cycle, a trend that reduces, other things equal, the aggregate output of innovators. This drop in productivity is particularly acute if innovators' raw ability is greatest when young.

JEL Codes: O31, J24, I2, B3, C14

Keywords: innovation, age, Nobel Prize, life-cycle productivity, education, economic growth, history of science, Kuhn

*I would like to thank Daron Acemoglu, Amy Finkelstein, Abhijit Banerjee, Leemore Dafny, James Dana, Ben Olken, Mike Mazzeo, Joel Mokyr, Sendhil Mullainathan, Scott Stern, Peter Temin, and Bruce Weinberg for helpful comments and suggestions. Alexander Karlan, Gowoon Lee, and Atanas Stoyanov provided excellent research assistance. All errors are my own.

†Kellogg School of Management and NBER. Contact: 2001 Sheridan Road, Room 628, Evanston, IL 60208. Email: bjones@kellogg.northwestern.edu.

Age is, of course, a fever chill
that every physicist must fear.
He's better dead than living still
when once he's past his thirtieth year.
– Paul Dirac, 1933 Nobel Laureate in Physics

1 Introduction

It is widely perceived that great innovations are the provenance of the young. The sentiments of Dirac expressed above have been shared by Einstein, von Neumann and many other eminent scientists and mathematicians (Zuckerman & Merton, 1973; Simonton, 1988). Empirical investigations of this view tend to support the idea that innovative activity is greater at younger ages, although great achievement before the age of 30 is not typical. Rather, a researcher's output tends to rise steeply in the 20's and 30's, peak in the late 30's or early 40's, and then trail off slowly through later years (Lehman, 1953; Simonton, 1991).

While many great insights do occur at younger ages, it is also clear that innovators spend a large number of their early years undertaking education.[1] Indeed, human capital investments dominate the early part of the innovator's life-cycle. Learning a subset of the skills, theories, and facts developed by prior generations seems a necessary ingredient to innovative activity. Newton acknowledged as much in his famous letter to Hooke, "If I have seen further it is by standing on ye sholders of Giants". Dirac and Einstein, who produced major contributions at the age of 26, first went through significant educational periods and then built directly on existing work. Dirac built on Heisenberg's uncertainty principle and Hamiltonian mechanics, while Einstein's early insights built on the work of Planck and Maxwell. Certainly, innovation would be a very difficult enterprise if every generation had to reinvent the wheel.

[1] Research in the psychology literature suggests that substantial training periods – ten years at minimum – are a prerequisite to expertise in many fields, from science to sports, music, medicine, and chess (see Ericsson and Lehmann 1996 for a review).

These two observations suggest an intriguing tradeoff. If innovators are especially productive when young, but education is an important preliminary input to innovation, then the opportunity cost of the time spent in education may be significant. Moreover, how innovators approach this tradeoff may change as the economy evolves. For example, accumulations of knowledge across generations may create increasing educational demands, so that expanding time costs of education delay the onset of active innovative careers. This possibility poses a problem for innovation as it reduces, ceteris paribus, the lifetime output of individual innovators, especially if their potential is greatest when young.

In this paper, I show that the great achievements in knowledge of the 20th Century occurred at later and later ages. The mean age at great achievement for both Nobel Prize winners and great technological inventors rose by about 6 years over the course of the 20th Century. This aging phenomenon appears to be substantially driven by declining innovative output in the early life-cycle. Moreover, the early life-cycle effects appear to be substantially explained by increases in training.

Section 2 presents the main fact: there has been a substantial increase in the age at great invention. This trend appears distinctive, as the age of great achievement in athletics has not changed. I introduce several hypotheses for the trend. In one type of hypothesis, the life-cycle productivity of innovators may have shifted. For example, increasing educational attainment may delay the onset of active innovative careers. Alternatively, innovator productivity may increase at more advanced ages due to improved health, effort or an increased role for experience. In another type of hypothesis, the upward age trend in the data could simply reflect underlying demographic shifts. Since the population has become substantially older with time, we are more likely to draw older innovators today than we were at the beginning of the 20th century. Put another way, if people lived shorter lives in the past, then innovators in the past will also appear younger.

Section 3 tests between these competing explanations and locates any specific shifts in life-cycle productivity. I find substantial shifts in life-cycle productivity beyond any demographic effect. Specifically, there has been a large upward trend in the age at which innovators begin their active careers. The estimates suggest that, on average, the great minds of the 20th Century typically became research active at age 23 at the start of the 20th Century, but only at age 31 at the end - an upward trend of 8 years. Meanwhile, there has been no compensating shift in the productivity of innovators beyond middle age.

Section 4 presents additional analysis to further understand the delayed start to the career. I first examine data on age at Ph.D. and show that Ph.D. age increases substantially over the 20th Century. I next harness World Wars I and II as natural experiments, testing the idea that training is a prerequisite for innovation and showing that interruptions to training must be "made up" after the war. Next, I investigate cross-field, cross-time variation and show that variations in Ph.D. age typically predict variations in the age-invention relationship. Collectively, these analyses suggest that training plays a key role in explaining the age-invention patterns.

Section 5 clarifies interpretations of the empirical patterns and considers their implications. I present a simple theory to examine the relationship between human capital investments and life-cycle productivity and show how accumulations of knowledge within fields can provide a consistent explanation for the set of facts. Further evidence from "ordinary" inventions underscores this perspective and also shows that the aging phenomenon extends broadly across the innovator population. Section 5 closes by clarifying specific implications of the empirical patterns for core issues in economic growth and the history of science. In particular, I show how contractions in the life-cycle of innovation can help explain the decline in innovative output per researcher seen over the 20th Century. Section 6 concludes.

2 Age and Great Achievement

This section presents benchmark facts about the age of individuals at the time of their great achievements. Two types of achievements are considered. The first group considers knowledge-based achievements: research that leads to Nobel Prizes in Physics, Chemistry, Medicine, and Economics, as well as technological achievements presented in almanacs of the history of technology. The second group considers athletic achievements based on world-record breaking events in track and field, as well as Most-Valuable Player awards in baseball. The athletes are included for comparison, allowing us to highlight what is distinctive about knowledge-based achievement.

The data set uses established sources to identify great achievements in knowledge. Nobel Prizes are determined by committees of experts and are given in principle for a distinct advance. The technological almanacs compile key advances in technology, by year, in several different categories such as electronics, energy, food & agriculture, materials, and tools & devices. The year (and therefore age) of great achievement is the year in which the key research was performed. For the technological almanacs, this is simply the year in which the achievement is listed. For Nobel Prizes, which are retrospective, the year of achievement was determined by consulting various biographical resources. The Data Appendix describes the data collection and sources in further detail.

As a first look at the data, Figure 1 presents ages at great innovation, considering all 20th Century observations together. Three features are of immediate note. First, there is a large variance in age. The largest mass of great innovations in knowledge came in the 30's (42%), but a substantial amount also came in the 40's (30%), and some 14% came beyond the age of 50. Second, there are no observations of great achievers before the age of 19. Dirac and Einstein prove quite unusual, as only 7% of the sample produced a great achievement at or before the age of 26. Third, the age distribution for the Nobel Prize winners and the great inventors, which come from independent sources, are extremely similar over the entire distributions. Only 7% of individuals in the data appear in both the Nobel Prize and great inventors data sets.

The most surprising aspect of these data, however, becomes apparent when we consider shifts in this age distribution over time. To start, I run the following regression:

$$a_i = \alpha + \beta t_i + \gamma X_f + \varepsilon_i \quad (1)$$

where a_i is the age of individual i at the time of the great achievement, t_i is the year of the great achievement, and X_f are fixed effects for the field of the achievement and the country of the individual's birth. Results of this regression are presented in Table 1. We see that the mean age at great achievement is trending upwards by 5 or 6 years per century. These trends are highly significant and are robust to field and country of birth controls. Indeed, the controls cause the time trend to strengthen, rising to about 8 years over the course of the 20th century. The strengthening effect of the controls on the trend suggests a compositional shift in great innovation towards fields and countries that favor the young.

These trends can be seen in greater distributional detail in Figure 2, which presents the raw data again but divides the 20th Century into three chronological periods: from 1900-1935, 1935-1960, and 1960 to the present. This figure combines all unique individuals in the Nobel Prize and great inventors data sets. Here we observe a general shift of the age distribution away from younger ages. There is a distinct drop in the presence of those in their 20's and an increased presence of those in later middle age.

One obvious hypothesis for this outward age shift is a shift in the life-cycle productivity of great minds. Given that the early part of an innovator's career is dominated by education, one natural reason for a decline in early innovative potential may be an increase in the time spent in training. More generally, there may be relative increases in the productivity of *older* innovators directly; for example, due to improved health or an increased role for experience.

But we must be careful in how we interpret the distributional shifts we see. An alternative hypothesis for the outward age shift is a simple demographic effect. If the underlying population of innovators is getting older, then older innovators will be more likely to produce substantial innovations, even if the relationship between age and innovative potential is fixed. The greater the ratio of 50-year-old innovators to 25-year-old innovators, the more likely the Nobel Prize winning invention or greatest technological insight to come from one of the 50-year-olds. Such demographic effects may be important: certainly, life-expectancy and the average age of the population have risen substantially over the 20th century.

One reduced-form way to test between these ideas is a difference-in-difference style analysis. If we view the scientific/technological innovators as a "treatment" group experiencing effects peculiar to knowledge-based careers, then we might profitably attempt a comparison with "control" groups that are claimed on a priori grounds to be immune to such knowledge effects. An obvious choice for a control group is great achievements in athletics. Age-achievement profiles in athletics will also be influenced to some extent by demographic effects, but athletes presumably do not face increasing training demands over time – the rules of their games are both straightforward and fixed. Figure 3 compares the underlying distribution of MVP winners in baseball before and after 1960, dividing the data in half. We see that the entire distribution appears essentially stationary. Such stationarity is also seen (in a similar, unreported figure) in the ages at which individuals break world records in various track and field events.

The substantial upward trend in the age of knowledge-based achievement is absent in physical achievement. However, comparing these achievements cannot be wholly satisfactory, partly because the age distribution of athletes favors the young to a degree that knowledge-based achievement does not. Shifts in the population density beyond middle age will therefore not influence the age distribution of athletic achievement, while they still might influence the age distribution of great achievements in knowledge. More generally, this comparison does not help us pinpoint any distinct shifts in life-cycle productivity for knowledge-based careers.

The following section develops a formal econometric model to identify specific shifts in innovation potential, controlling for demographic effects. With this econometric model, we begin to open the black-box of the age-invention relationship by asking two questions explicitly. First, is the upward trend in the age of great achievement simply a demographic effect, or is it driven by shifts in innovator's life-cycle productivity? Second, if life-cycle productivity is shifting over time, is this due to effects at the beginning of the life-cycle, the end of the life-cycle, or both?

3 Life-Cycle Productivity

This section presents an econometric model to define the probability that witnessed innovations are produced by innovators at particular ages. Empirical analysis follows, using this model to determine sources of the upward trend in the age of great achievement.

Given a great innovation, the probability that this innovation was produced by an individual of age a will depend on two things. First, it will depend on the relative innovation potential of innovators in different cohorts. For example, according to Dirac, a physicist below the age of 30 has more innovative potential on average than one who is older. Second, it will depend on the density of innovators of various ages. If a population is full of 50-year-old researchers but has very few 20-year-old researchers, then the likelihood a particular innovation came from a 20-year-old is low, even if young innovators have good ideas.

Define the age distribution of the population at time t as $p_a(t)$. Next define $\bar{x}_a(t)$ as the average innovation potential of a given cohort at time t. The probability a given innovation comes from an innovator of age a is then

$$\Pr(a|t) = \frac{p_a(t)\bar{x}_a(t)}{\sum_{\{a \in A\}} p_a(t)\bar{x}_a(t)} \tag{2}$$

This expression is derived formally in the appendix by aggregating innovative potential over individuals in the population, but it should be intuitive. The probability a given innovation comes from a person of age a is just the relative innovation potential across cohorts weighted by the population age density, or, equivalently, the population age density weighted by the innovation potential.

Several useful points can now be made. First, shifts in the age distribution of great innovation, as seen in Figure 2, must be driven either by shifts in the population age distribution, $p_a(t)$, or by shifts in the average innovation potential of various age groups, $\bar{x}_a(t)$. Second, the stochastic process represented in equation (2) can produce innovators with a large variance in age, as demonstrated in Figure 1. Third, any presumption that the innovators' upward age trends are driven mechanically by increasing life expectancy may be misleading if the innovation potential, $\bar{x}_a$, of those in their later years is low – if only because people retire. Finally, it is worth noting that this stochastic model makes few assumptions. While we will make further assumptions in how we define $p_a(t)$ and $\bar{x}_a(t)$, the model to this point is quite general.

Equation (2) is the central vehicle for the maximum likelihood estimation to follow. In particular, given data for the population distribution, $p_a(t)$, and a series of year-age observations for great achievements, we can use (2) to test hypotheses about the shape of innovation potential, $\bar{x}_a(t)$. Before continuing to the estimation, it remains to develop an explicit model of $\bar{x}_a(t)$ and how it may shift over time. This sub-model is presented in the next section.

3.1 A Model of Life-Cycle Productivity

In this section we add parametric structure to the definition of life-cycle innovation potential, $\bar{x}_a$. In choosing an appropriate modeling strategy, it is helpful to first consider the existing empirical literature on creative careers, which suggests the following general pattern (e.g., Lehman 1953, Bloom 1985, Simonton 1991, Stephan and Levin 1993). First, the life-cycle begins with a period of full-time training in which there is no substantive creative output. Second, there is a rapid rise in output, following an S-curve, to a peak in the late 30's or early 40's. Third, innovative output declines slowly through later years, following a declining S-curve. While laboratory experiments do suggest that creative thinking becomes more difficult with age (e.g. Reese et al, 2001), the decline in innovative output at later ages may largely be due to declining effort, which a range of sociological, psychological, institutional, and economic theories have been variously proposed to explain (see Simonton 1996 for a review).

Given this pattern, consider the following simple model. Supressing time subscripts for the moment, we write

$$\bar{x}_a = L_1(a)L_2(a) \tag{3}$$

where $L_1(a)$ captures early life-cycle effects and $L_2(a)$ captures late life-cycle effects. (The appendix provides a derivation of this model as the aggregation of individual innovation potential.) We imagine that $L_1(a)$ is close to zero in the earliest part of the life-cycle and then rises once innovators complete their training and become active researchers. To estimate $L_1(a)$, we assume a logistic specification

$$L_1(a) = \frac{1}{1 + e^{-(a-\mu)/\omega}} \tag{4}$$

where μ is the average age at which the career begins and ω is a variance parameter. A logistic specification seems reasonable as it is parametrically simple, flexible, and captures the "S" shape one sees in early life-cycle output.[2] Figure 4 presents a graph to clarify this logistic specification and the meaning of the parameters.

To estimate $L_2(a)$, we assume a second logistic curve with parameters θ and ρ,

$$L_2(a) = 1 - \frac{1}{1 + e^{-(a-\theta)/\rho}} \tag{5}$$

This reverse "S" curve appears reasonable, as it can capture the initially slow decline in output in later middle age, followed by the more rapid decline and then tailing off of output into old age, as documented in the literature noted above.

[2] Another natural option is a normal distribution, a variation that has no substantive effect on the results.

With equation (3) and its sub-components (4) and (5), we now have a model for innovation potential over the life-cycle. We can estimate this model to determine how the propensity to produce great achievements in knowledge changes with age. Moreover, by articulating specific underlying models for both the "front end" of the life-cycle, (4), and the "back end" of the life-cycle, (5), we can ask not only whether innovation potential has been shifting over time but, more specifically, whether any shifts are coming from the early years of life, the late years of life, or both.

A motivational question in this paper is whether μ, the mean age at which the active career begins, is changing over time. Shifts in this mean over time can be generally modeled by a polynomial expansion,

$$\mu(t) = \mu_0 + \mu_1 t + \mu_2 t^2 + ... \tag{6}$$

Shifts in the variance parameter can be modeled similarly. The main estimation below will allow for a linear trend in $\mu(t)$ and a fixed variance parameter, ω; more general specifications will also be considered as robustness checks.

As with the beginning of the innovative career, we can further allow for shifts in innovation potential at the end of the career,

$$\theta(t) = \theta_0 + \theta_1 t + \theta_2 t^2 + ... \tag{7}$$

For example, as noted above, shifts that increasingly favor experience over raw ability may increase later life innovation potential. Meanwhile, improved health technology may lead to clearer thinking and/or increased physical stamina, while, alternatively, rising incomes could encourage earlier retirement and a decline in average innovation potential among older innovators. In the estimation I will constrain $L_2(30) > 0.9$ to ensure that (5) and movements in it are describing effects later in life, which will make the results more transparent to interpret. This strategy will help us to substantially limit theories for the increased age at great innovation over the 20th Century.

Taken together, equation (2) and the sub-model of innovation potential given by equations (3)-(7) produce a stochastic model that integrates demographic effects with a model of knowledge accumulation.[3]. The following subsection discusses the data used to estimate $p_a(t)$. We then present the central results.

3.2 Population Data

The great innovators come from many different countries and are therefore drawn from populations with differing age distributions. Data on these age distributions are difficult to find for many countries, particularly over the time-frame of the entire 20th century. For this reason, the estimation will focus on the American subset of great innovators. The American innovators show a similar trend in mean age at great achievement as the larger group and provide a significant number of observations on their own.[4]

The population age densities are calculated from large micro-samples of the U.S. census. With these micro-samples, it is possible to determine not only the age distribution for (i) the entire national population, but also the distribution for subgroups of (ii) active workers and (iii) professional scientists and engineers. The scientist and engineer data are appealing as they may capture a closer approximation of the relevant age distribution of innovators. However, the sample sizes are small in early census years, and the occupational codes in the census are not entirely consistent across time, raising concerns that shifts in the age distribution for this sub-group may partly be an artifact of shifting classifications. The maximum likelihood model will be estimated using each of these population data sets. As we will see, the estimates are quite insensitive to the choice of population. See the Data Appendix for further discussion of these census data and the construction

[3] The log-likelihood function is: $\sum_i \log\left(\frac{p_{a_i}(t_i)\frac{1}{1+e^{-(a_i-\mu(t))/\omega(t)}}\left(1-\frac{1}{1+e^{-(a_i-\theta(t))/\rho(t)}}\right)}{\sum_{\{a_i\in A\}} p_{a_i}(t_i)\frac{1}{1+e^{-(a_i-\mu(t))/\omega(t)}}\left(1-\frac{1}{1+e^{-(a_i-\theta(t))/\rho(t)}}\right)}\right)$

[4] There are 294 American-born great innovators. The trend in age at great achievement is 8.24 years/century with a standard error of 2.58 years/century.

of the science and engineers sub-sample.

3.3 Results

Table 2 presents the maximum likelihood estimates for shifts in innovation potential. Columns (1) through (3) allow for linear trends in the mean parameters, which we take as the main specification.

There are two striking results. First, there has been a large shift in life-cycle innovation potential, even when controlling for an aging population. Second, the shift in innovation potential is felt entirely at the beginning of the life cycle. In particular, we see that the mean age at which innovators begin making active contributions has increased by about 8 years over the course of the 20th Century, rising from a mean age of about 23 in 1900 to approximately 31 in the year 2000. These results are robust to the choice of population data. Meanwhile, there is little shift in innovation potential in middle age or beyond. Depending on the specification, estimates show at most modest and in all cases highly insignificant movements. Interestingly, this stationarity implies that demographic effects have driven the rising density of innovators beyond middle age that was seen in Figure 2.

Figure 5 compares the estimated life-cycle curves for the year 1900 and the year 2000, using specification (3). We see that the peak ability to produce great achievements in knowledge came around age 30 in 1900 but shifted to nearly age 40 by the end of the century. An interesting aspect of this graph is the suggestion that, other things equal, lifetime innovation potential has declined. This point will be further discussed in Section 5.

One may also now decompose the trends of Section 2 into demographic and productivity components. Holding population distribution fixed using 1950 data, the productivity shift in Table 3 implies an approximately 5 year increase in the mean age of great achievement. Meanwhile, holding innovation potential fixed at its 1950 estimate, the aging population suggests an approximately 3 year increase in mean age. Hence productivity shifts account for about 60% of the 8-year age trend seen in Section 2, while the aging population captures 40%.[5]

These basic results are robust to a number of alternative specifications. The specifications in Table 3 allow for additional linear trends in the variance parameters of the logistic distributions. We see that the only statistically significant movement in innovation potential continues to come at the beginning of the life cycle, in the mean age at which the career begins. Other functional forms have also been examined, including those that do not assume logistic distributions. Regardless of the functional form, I find that the significant trends appear only in the beginning of the life-cycle, showing an increasing age at which innovators begin innovating.[6]

Columns (4) and (5) of Table 2 further consider the U.S. Nobel Prize and U.S. great inventor datasets separately, showing similar results in each case. While the Nobel Prize is in principle given for distinct achievements, we might be concerned that other criteria affect the selection, and that these criteria have shifted over time to favor older innovators.[7] Possible selection concerns regarding Nobel Prizes are unlikely to be important here, however, mainly because the great inventor data set, which simply lists the great technological achievements in each given year, appears more immune to these kinds of selection biases and yet, on every dimension, has produced similar results.[8]

[5] Moreover, much of the demographic effect comes in the first 2/3rds of the century; more recently, the baby-boom significantly attenuated the aging of the working population and even reversed it in the 1970s and 1980s.

[6] The logistic functions are appealing for the reasons discussed in Section 3.1. The estimation methodology can, however, also employ polynomial or piecewise linear functions to describe sequential pieces of the life cycle. Estimations become noisier as the number of estimated parameters increases, but they produce similar basic results: an increase, prior to middle age, in the age at which innovators begin producing great ideas.

[7] For example, an increasing bias towards lifetime achievement could have this effect.

[8] The Nobel Prize and great inventors data sets have extremely similar age distributions (Figure 1) and extremely similar mean trends (Table 1). Table 2 shows that the structural trends are similar for both groups when they are estimated independently; the coefficients for the great inventors are the same as for the whole, and the standard errors rise slightly as would be expected given the smaller sample size. These common patterns suggest common forces rather than idiosyncratic selection effects. Finally, the results of Tables 2 and 3 show shifts in innovation potential at the *beginning* of the life-cycle and not at the end, which is not consistent with selection stories based on longevity or increasing favoritism for lifetime achievement.

4 Inside the Early Life-Cycle

Age at great invention has trended upwards by approximately 6 years over the course of the 20th Century. This trend is not simply due to an aging population but reflects a substantial change in the life-cycle productivity of innovators. Furthermore, the maximum likelihood estimates focus interpretations on effects limited to the young. Explanations must confront not a general aging effect but a specific, substantial delay at the beginning of the life-cycle.[9]

A natural and intriguing hypothesis for the rising delay in the early life-cycle is the possibility that training time has increased. A viewpoint that emphasizes training would build on two claims. First, that training is an important preliminary input to the innovative career. Second, that variations in training duration can help explain the age-invention relationship.

This section focuses on the early life-cycle and the role of training to further open up the black box of age and invention. I undertake three analyses. The first analysis looks directly at evidence from Ph.D. age and shows that Ph.D. age increases substantially over the 20th Century. The second analysis harnesses world wars, as exogenous interruptions to the young career, to test the basic idea that training is an important preliminary input to innovation. I show that, while the world wars do not explain the 20th century's age trend, they do indicate the unavoidable nature of training: lost years of training appear to be "made up" after the war. The final analysis explores cross-field, cross-time variation. I show that variations in training duration predict variations in age at great invention, and I close by discussing a perspective in which shifts in foundational knowledge explain major training and achievement age patterns within fields and over time.

[9] Theories that focus on productivity in the later life-cycle, such as improved health effects, find little support. Theories that suggest delays in innovation at both young and old ages will also have trouble explaining the specific empirical patterns we see. For example, research on creative careers in the arts (Galenson & Weinberg 2001; Galenson 2004a; Galenson 2004b) has suggested a useful distinction between "conceptual" innovation and "experimental" innovation, where the former favors the young and the latter favors the old – often the very old. However, these important ideas are not wholly satisfactory here because an increasing experimental bias would presumably be felt to a large degree at older ages.

4.1 The Age at Highest Degree

Given the increasing delay at the beginning of the life-cycle, an obvious question is whether this delay is reflected in longer periods of formal education.[10] While the Ph.D. is an institution that only approximately captures the end of the training phase and the beginning of the primary research phase, it is also the most obvious delimiter between these phases. I consider here the basic trend.

For 93% of the Nobel Prize winners, it was possible to determine the age and location for the highest degree. In 96% of these cases, the highest degree was a doctorate. I analyze trends in the age by running the following regression:

$$a_i^D = \alpha + \beta t_i^D + \gamma X_f + \varepsilon_i \tag{8}$$

where a_i^D is the age of individual i at the time of their highest degree, t_i^D is the year of the highest degree, and X_f are fixed effects for the country of the degree and the field of the ultimate achievement.

The results are presented in Table 4. We see that Nobel Prize winners complete their formal education at substantially older ages today than they did a century ago. There is an upward age trend of approximately 4 years per century, and the trend is robust across specifications. This result suggests that training duration may be intimately related to the drop in innovative output in the early life-cycle.[11] As we will see below, patterns in the age of highest degree also inform

[10] Indeed, several studies have documented upward trends in educational attainment among the general population of scientists. For example, the age at which individuals complete their doctorates rose generally across all major fields in a study of the 1967-1986 period, with the increase explained by longer periods in the doctoral program (National Research Council, 1990). The duration of doctorates as well as the frequency and duration of post-doctorates has been rising across the life-sciences since the 1960s (Tilghman et al, 1998). A study of electrical engineering over the course of the 20th century details a long-standing upward trend in educational attainment, from an intitial propensity for bachelor degrees as the educational capstone to a world where Ph.D.'s are common (Terman, 1998).

[11] Interestingly, while the increase in training age is large, it accounts for only half the shift seen in the maximum likelihood estimates, although it is within those estimates' confidence intervals. Institutional variations in Ph.D. requirements may complicate further interpretations. For example, the country fixed effects in (8) are jointly significant with a p-value of less than .0001. This suggests that variations in degree requirements differ across countries; institutional variations over time are then likely as well. The well-known rise of post-doctorates (e.g. Tilghman et al, 1998) and/or increased "on-the-job training" could suggest further extension of the training phase in ways not captured by charting ages at Ph.D..

substantially more detailed variation in the data.

4.2 War Interruptions

World Wars I and II created interruptions in many careers. For certain cohorts, the interruption of the war was felt largely in the training phase. This provides a natural experiment to investigate the role of training in the early life-cycle. I consider two types of analysis, one that relies on educational data and one that does not.

First, consider that in every year there are people who have completed their undergraduate degree but not their graduate degree. Every person is at some point between degrees, so by drawing any year at random we draw a sample of people with similar innate characteristics, on average. We can then ask whether those individuals who happened to be between degrees at the outset of world war, in 1914 and 1939, as opposed to being between degrees in other years, experienced unusual delays in completing their training and in their ensuing innovative careers. For 68% of the Nobel Prize winners, it was possible to collect the year of the undergraduate degree and hence identify individuals who are between degrees. I run regressions of the form

$$y_i = \alpha + \gamma_1 WW1 + \gamma_2 WW2 + \beta t_i + \gamma X_f + \varepsilon_i \tag{9}$$

where y_i is the outcome variable of interest: either the age at highest degree, the number of years between the undergraduate and graduate degree, or the age at great achievement. The variables $WW1$ and $WW2$ are dummies equal to 1 if the individual happened to be between degrees at the outset of the indicated war. The control t_i captures background trends in the dependent variables, and X_f includes field and country fixed effects and dummies for cases where educational data is not observed.

The results of these regressions are presented in the left panel of Table 5. We see that both world wars resulted in a 2 year increase in the age at Ph.D. for those individuals who happened to be caught between degrees. Related, there is 2 to 3 year increase in the number of years between the undergraduate and graduate degree. These both suggest that interruptions to training must be "made up". Table 5 shows that these early life cycle delays can be further associated with increased age at great achievement, by 2-3 years for World War II although there is little effect for World War I. Given that innovation potential remains high through middle age, it is not clear that a small number of observations with early life-cycle interruptions will show an increase in the mean age of great innovation.[12] One expects, more precisely, a decline in innovative potential at younger ages, which is shown next.

Figure 6 presents the percentage of great achievements produced by those aged 25 to 30 in each five year period over the 20th century. This figure uses the full sample of great minds, not just those for whom educational data is available. For comparison, the percentage of the U.S. workforce between these ages is also shown. Interestingly, we see a particularly sharp decline in the proportion of innovators aged 25-30 in the five years *after* the world wars (1920-1925 and 1945-1950). These cohorts were aged 20-25 during the wars - during the typical training phase - yet these cohorts appear extremely unproductive *after* the war, relative to other innovators in the immediate postwar and relative to the typical productivity of this age group in immediately previous and subsequent intervals.

To further investigate this pattern, I run the following Probit model for the great achievement data

$$\Pr[25 <= a_i < 30] = \alpha + \psi_0 treatment_i + \psi_1 control_i + \beta t_i + \gamma X_f \tag{10}$$

where *treatment* is a dummy equal to 1 for those aged at least 20 at some point during the war and never more than 25, while control additionally captures those aged 15-19.[13] Other controls

[12] The weak World War I result in particular may reflect low power; since there are only 10 individuals in the data with $UGONLY\_WW1 = 1$ (there are 66 observations in the World War II case).

[13] Control cohorts are born 1893-1903 (World War 1 case) and 1920-1930 (World War 2 case). Treatment cohorts are born 1893-1898 (World War 1) and 1920-1925 (World War 2). World War I was primarily fought 1914-1918 and

are as above. The results show that individuals who experience war between the ages of 20 and 25 had an unusually low probability of innovating in the 25-30 age range – even though the wars were over. This pattern holds independently for both world wars. A natural explanation is again that interruptions in the training phase had to be "made up". This analysis is based on the output measure - great achievement - rather than education data, yet the conclusions are substantially the same. Great minds do not magically arrive at high innovation potential at a certain age, but rather their behavior in their early life-cycle informs their ensuing innovative output. In particular, interruptions during the training phase create delays to their education and their achievements, suggesting that training is an important preliminary input to the innovative career.[14]

4.3 Knowledge Accumulation and Revolution

We can go further in understanding any training-achievement nexus by harnessing additional variation in the data. Figure 7 plots the evolution of age at great invention separately for the four Nobel sub-fields (dark lines, left axis). These are non-parametric regressions so that the patterns are seen without imposing a functional form. The age at Ph.D. is separately plotted (grey lines, right axis), as are 95% confidence intervals.[15]

We see that Ph.D. and achievement age tend to follow remarkably similar dynamics within fields. The shared dynamics between Ph.D. age and achievement age are most apparent in the hard sciences – physics, chemistry, and medicine – and less so in economics, although this case is obscured by outliers.[16] Most strikingly, both achievement and Ph.D. age in physics experienced a

World War II late 1939-1945.

[14]Note also that isolating the influence of world wars on these groups does not change the overall 20th century trends. While the wars lead to interruptions in training and substantially reduce early life-cycle innovation potential, we see in Table 5 that the background trends in age at great achievement and age at Ph.D. are essentially unchanged. Figure 6 further underscores this point and, more broadly, the maximum likelihood estimates of Section 3. The figure shows a continuing drop in innovative output among the young over the whole 20th century, a drop far greater than shifts in population density alone, or the abberations of wars, would suggest.

[15]These plots are the results of Fan regressions with a quartic kernel, 25% bandwidth, and bootstrapped standard errors (Fan 1992; Deaton 1997).

[16]The movement in mean Ph.D. age for Economics is less predictive, although the confidence intervals are wide. There are only 53 observations for economics, which limits the inference, and the sharp rise to 1960 is driven largely

unique decline in the early 20th century. This unusual feature, beyond reinforcing the relationship between training and achievement age, may also serve to inform basic theories for the underlying dynamics and differences across fields.

An intriguing side-effect of innovation is the possibility that new ideas impose an increasing educational burden on future innovators. If the set of foundational ideas expands, training time may expand, making innovators less productive in their early life-cycle. On the other hand, progress in science need not require expansions in foundational ideas. New ideas sometimes serve not as extensions or refinements but rather as revolutions, leading to contractions in the knowledge space that may cause training requirements to decline. Whether scientific progress is fundamentally cumulative or revolutionary in nature is an empirical question - and one much debated by historians of science. Thomas Kuhn distinguished between periods of "normal" science (accumulation) and periods of "paradigm" shifts (revolution), with early 20th Century physics as his quintessential example of the latter (Kuhn, 1962).

Jones and Weinberg (2006) consider the U-shaped age relationship in physics through the lens of accumulation and revolution. Drawing on that analysis, it is clear that the advent of quantum mechanics in physics, a revolution that reached into virtually all the field's sub-specialties, is coincident with the unusual behavior in the age data. Many historians chart the specific period from 1900-1927 as the time when the entire worldview of physics changed (e.g. Kuhn, 1962; Jammer, 1966; Galison et al., 2002).[17] From a training point of view, physicists found themselves in the

by a few significant outliers (Allais, Coase, and Stone, who receive their Ph.D.s at ages 38, 41, and 44 respectively, long after their first academic appointments and years of successful research). These outliers mask a continuing decline in the completion of formal education at young ages. Prior to 1950, 32% of eventual Nobel Prize winners in Economics completed their highest degree prior to age 25. After 1950, only 7% completed their highest degree by age 25. Similar, large declines in Ph.D. completion by the very young are also seen in Physics, Chemistry, and Medicine.

[17] Starting with Planck's idea in 1900 that radiation comes in discrete energy packets (quanta), the work of Einstein, Bohr, Compton, Hertz and many others followed, breaking down the classical world of physics and reaching firm footing only with the advent of a consistent "quantum mechanics", developed by Schrodinger, Heisenberg and others in the mid to late 1920s.

early 1900s wrestling with a new, limited set of empirical puzzles and the failure of existing theory, allowing young minds to achieve the research frontier relatively easily. The firm establishment of quantum mechanics in the late 1920s then led back towards normal science, a long period of accumulation and refinement, that has continued since.[18] In the data, the decline in age during the early 20th century among physicists reaches a minimum in the late 1920s, coincident with the establishment of quantum mechanics, and then rises thereafter. Similar patterns occur in Ph.D. age. The decline in age appears closely associated with this revolution of knowledge, which was distinct to physics, and is not seen in other fields.

In sum, training dynamics appear to usefully predict variations in the age-invention relationship. Digging deeper, the distinction between knowledge accumulation and revolution can provide a useful lens through which to view major training and age-invention patterns in the data, an interpretation that is further explored below.

5 Interpretations and Implications

The evidence in this paper points to early-life cycle effects and a particular role for training in understanding the age-invention relationship. In this section, I explore more formal reasoning for innovators' training decisions. I clarify the potential explanatory power of several hypotheses and close by considering implications of the patterns uncovered in this paper.

Consider the following simple model. An innovator is born with no knowledge but endowed with time. Innovators can invest in training, but knowledge acquisition delays the active production of new ideas. Innovators compare the return to active production with the return to further training, whatever the benefits that training brings.

A reasonable specification, especially for highly motivated innovators, is that educational attainment is chosen to maximize one's lifetime research contribution. In particular, the choice

[18] While controversial, some argue that string theory is an incipient revolution at this juncture.

problem is:[19]

$$\max_E \int_E^T f(E)g(a)da$$

where $f(E)$ represents the value of education to their innovative output, and $g(a) > 0$ represents the individual's natural ability as a function of age.[20] Individuals spend some number of years, E, focused on education during which time they do not innovate, followed by a career of innovation until they die at time T. The amount of education influences their ultimate productivity, where $f'(E) > 0$ and $f(0) = 0$.

The first order condition for this problem is:

$$f'(E) \int_E^T g(a)da = f(E)g(E) \tag{11}$$

which clarifies the central tradeoff. Greater education brings a benefit: the incremental effect on innovative output, $f'(E)$, weighted by the innovative ability that remains over your lifetime. But it also brings an opportunity cost, $f(E)g(E)$, the current innovation potential foregone. Figure 8 presents the optimization condition (11), with the terms rearranged. I label the upward sloping curve the "longevity curve" and the downward sloping curve the "training curve".[21] Increasing training time, E^*, is an endogenous response to rightward shifts in either curve.

Consider the implications of expansions in foundational knowledge. If knowledge accumulates across generations, the training curve can shift rightward. This is a natural outcome if (a) there is a set of preliminary skills one must master to reach the frontier in a field, and (b) the set of these skills expands. For example, imagine that innovators need to know A, B, C to attack D, the frontier. Following success at D, however, ensuing cohorts may need to know A, B, C, D, to

[19] More generally, this objective function captures cases where utility is defined to maximize fame or income, so long as fame and income are monotonic functions of lifetime innovative output.

[20] One can also incorporate time discounting in the $g(a)$ function.

[21] The longevity curve has some finite value at age 0 and approaches ∞ as $E \to T$, since $\lim_{E\to T} \int_E^T g(a)da = 0$. The educational return curve has some finite value at age T and approaches ∞ as $E \to 0$, since $f(0) = 0$ and therefore $\lim_{E\to 0} f'(E)/f(E)$ is unbounded. Hence there is an interior solution to this problem. I have drawn these curves as monotonic in E, so that the curves have a single crossing property, but this is not necessarily the case under general functional specifications.

attack the new frontier.[22] Foundational skills are complements, with the mastery of many needed to become capable of frontier research. In its simplest form, one imagines f as a step function with a step up after E years of training.

We may also imagine that the longevity curve shifts rightward. Increases in life expectancy (T) would have this effect.[23] Longer life expectancy can lead endogenously to increased training time by providing longer lives over which to reap the fruit of formal training. There are several reasons to believe, however, that such a mechanism may not provide an adequate explanation for the empirical patterns. First, mean life expectancy at age 10 was already greater than 60 in 1900, while it is clear from Sections 2 and 3 that innovation potential is modest beyond 60, so that adding years of life beyond this age would have at most mild effects on the optimization.[24] Related, even modest discounting would substantially limit the effect of gains felt 35+ years beyond the end of training on the marginal training decision. Next, common life expectancy changes cannot explain the unique cross-field and cross-time variation explored in Section 4, such as the unique behavior of physics. Moreover, Figure 7 suggests, if anything, accelerating age trends after the second world war, which is hard to explain with increased longevity, where post-war gains have slowed.

Another set of explanations involve institutional or sociological effects on the training curve. A plausible story might involve signaling. If establishing a reputation is a prerequisite for grant-based research, and research has become more grant-based (expensive) over time, then extensions of formal training or apprenticeship may serve to signal reputation - a different interpretation of the educational return curve. While this force may be operating, it also suffers as a general explanation;

[22] A revolution would result if overcoming D involves overturning foundational knowledge as well, so the path to the frontier changes, e.g., to A, B', D, with B' replacing the old B, C.

[23] Increases in innovation potential beyond middle age, an increase in $g(a)$, would also have this effect. The evidence from Section 3 however, indicates that $g(a)$ is not shifting after middle age.

[24] This life expectancy data is for white males in the United States. (Source: Department of Health and Human Services, National Center for Health Statistics; National Vital Statistics Reports, vol 53., no. 6, Nov. 10, 2004.) The life-expectancy of innovators, who have several advantages, would likely be higher still. In fact, the age at death for great innovators in the sample who were born between 1900 and 1910 averages 80.

for example, Figure 7 shows a substantial age increase at great achievement among the economists alone, and these prizewinners have had little need for large grants.

Other forces might also be considered within this framework, none of which are mutually exclusive and many may be operating.[25] At the same time, reasoning about shifts in foundational knowledge appears to provide a parsimonious explanation for the major patterns in this paper. A difficulty in directly establishing this thesis is the difficulty in directly measuring the stock of foundational knowledge - the "distance to the frontier". But we can go further by considering additional, indirect evidence. To see this, we further articulate the training decision.

If we imagine that the innovators must reach the frontier to do innovative work, the innovator still has a decision over the breadth of expertise along the frontier. To fix ideas simply, let

$$E = bD$$

where D measures distance to the frontier and b measures the innovator's breadth along the frontier. Total training time is then determined in part by the capacity for specialization. For example, the chemist may choose to come a frontier expert in the synthesis of metal alloys (narrower b), or both alloys and organics (wider b).[26] This specialization margin presents a useful empirical application. Inferences about knowledge accumulation, D, based only on investigations of training time, E, may be clouded by other possible forces as discussed above. But we might infer knowledge accumulation more definitively by observing E in combination with some measure of breadth of expertise, b. That is, $D = E/b$. If people spend longer in training (more E) and yet come out the

[25] As a final example, one may imagine declining educational efficiency, shifting the training curve rightward. Such a decline in training efficiency would have to be very large to explain the estimated 8 year delay, from age 23 to 31, in achieving high innovation potential, and one might imagine that educational efficiency increases with technology, rather than decreases, suggesting some skepticism for this particular point of view. The biographies of Nobel prize winners further suggest a degree of focus that is not commensurate with slow or undirected training.

[26] The capacity for specialization will likely be imperfect. For example, the chemist may specialize more or less on certain types of synthesis but regardless must understand theories of valence and molecular structure. The knowledge space is thus a mix of common foundational ideas and more specialized ideas. When the stock of knowledge grows, we would typically imagine that innovators would respond on both dimensions, partly by increasing their training time and partly by increasing specialization (though the following argument is empirical and does not require this assumption).

other end more specialized (less b), then the distance to the frontier has increased.

This reasoning is explored in Jones (2005), which studies "ordinary" inventors, looking at all U.S. patents in the 1975-2000 period. There are two key results. First, the age at first patent is rising at a rate of 6 years/century. Age at first patent provides an outcome-based measure to delimit the training and research phases. Remarkably, this estimate for the extending training period is extremely similar to the trends in this paper. Second, proxy measures for specialization show increased specialization across the full range of technological fields. One proxy measure is research collaboration in patenting - measured as team size - which is increasing at over 10% per decade.[27] A more direct measure of specialization considers the probability that an individual switches technological areas between consecutive patents. Jones (2005) shows that the probability of switching technological areas is substantially declining with time. These analyses indicate that training time, E, is rising, while measures of breadth, b, are simultaneously declining. It is then difficult to escape the conclusion that the distance to the knowledge frontier is rising. This evidence also acts to confirm - with "ordinary" invention - the rising age pattern found among the great minds.

5.1 Implications

Shifts in life-cycle research productivity can have diverse implications, from the efficient targeting of grants to the design of tenure processes and the timing of child rearing. Here I will emphasize two aggregate implications, for core issues in economic growth and scientific progress, that are suggested by the particular life-cycle shifts identified in this paper.

First, other things equal, the shorter the period that innovators spend innovating, the less their output as individuals. If innovation is central to technological progress, then forces that reduce the length of active innovative careers will reduce the rate of technological progress. This effect will be

[27] Large and general upward trends in research collaboration are also found in journal publications (e.g. Adams et al, 2004).

particularly strong if innovators do their best work when they are young. In fact, aggregate data patterns, much debated in the growth literature, have noted long-standing declines in the per-capita output of R&D workers, both in terms of patent counts and productivity growth (Machlup 1962; Evenson, 1991; Jones 1995a; Kortum, 1997). Simple calculations from aggregate data suggest that the typical R&D worker contributes approximately 30% as much to aggregate productivity gains today as she did at the opening of the 20th Century.[28] This paper provides micro-evidence that can explain part of that trend. Other things equal, the estimates of Section 3 indicate a 30% drop in the lifetime innovation potential over the century, or nearly half of the overall decline in individual research productivity.[29,30]

Second, the facts in this paper can also inform basic debates about the nature of scientific progress. A core question in the history of science is whether scientific progress happens primarily through accumulation and refinement of ideas or through radical, "Kuhnian" revolution. These debates are traditionally enjoined through historical argument, such as Kuhn's seminal analysis of physics. The age data in this paper can provide, alternatively, a data-driven test. If the aging phenomena detailed in this paper suggest, as discussed above, accumulations of knowledge, then

[28] Combining Machlup's data on growth in knowledge producing occupations for 1900-1959 (Machlup 1962, Table X-4) with similar NSF data for 1959-1999 (National Science Foundation, 2005), we see that the total number of knowledge-producing workers in the United States has increased by a factor of approximately 19. Meanwhile, the U.S. per-capita income growth rate, which proxies for productivity growth over the long-run, suggests a 6-fold increase in productivity levels (based on a steady growth rate of 1.8%; see Jones 1995b). The average rate at which individual R&D workers contribute to productivity growth is $\dot{A}/L_R$, or gA/L_R , where A is aggregate productivity, g is the productivity growth rate, and L_R is the aggregate number of R&D workers. The average contribution of the individual R&D worker in the year 2000 is then a fraction $\left(A^{2000}/A^{1900}\right)/(L_R^{2000}/L_R^{1900}) = 6/19$ (32%) of what it was in 1900.

[29] This paper estimates the relative innovation potential across age groups, so that forces that enhance or reduce the impact of all innovators, regardless of age, are not captured. Other influences, on top of delays at the beginning of the life-cycle, may therefore help to explain further portions of the declining trend in the average contributions of innovators. Suggested mechanisms include innovation exhaustion or "fishing out" stories (e.g. Evenson, 1991; Kortum, 1997), as well as narrowing expertise and innovative capacity as an endgenous response to an increased educational burden. Jones (2005) provides theory and empirical support for this latter mechanism.

[30] Note that these aggregate implications, stemming from the career contraction, come regardless of the particular cause of the early life-cycle delay. If purely institutional effects such as signaling were responsible, then there may be nothing inevitable in the career contraction going forward, and the contraction may be highly elastic to institutional reform. Meanwhile, there is considerable evidence that shifts in foundational knowledge can explain several dimensions of the age patterns, and other patterns. In this case, the delay in the early life-cycle may be an efficient response to a by-product of technological progress - the accumulation of foundational knowledge.

Kuhnian revolutions appear rare.

6 Conclusions

Great minds produce their greatest insights at substantially older ages today than they did a century ago. This upward age trend is not due simply to an aging population, but comes from a substantial decline in the innovative output of younger innovators. Meanwhile, there is no compensatory expansion of innovative output at later ages. Innovators are the engines of technological change and, other things equal, the less time an innovator spends successfully innovating, the less her lifetime output. The estimates point to a 30% decline in life-cycle innovation potential over the 20th Century.

This paper further explores the role of training in understanding the early life-cycle dynamics, investigating evidence from world wars, age at Ph.D., and cross-field, cross-time variation in training and achievement ages. These analyses further unpack the black box of the age-invention relationship and point towards the training phase as a key explanation for the trends we see. Yet the economics literature has focused little on the human capital investments of innovators. Given that innovators spend some of their youngest and potentially brightest years undertaking educational investments, understanding the tradeoffs at the beginning of the life-cycle may be first-order for understanding the ultimate output of these individuals. Certainly, great innovation is less and less the provenance of the young.

7 Appendix

This Appendix derives the econometric model of Section 3 as an aggregation across the stochastic behavior of individuals.

Consider a population N. Given a witnessed innovation, the probability the innovation was produced by an individual i is defined by:

$$\Pr(i) = \frac{x_i}{\sum_{\{i \in N\}} x_i}$$

where x_i represents the innovation potential of person i. Innovation potential measures the relative innovative strength of an individual.[31]

For estimation, consider the model in terms of cohorts of equally-aged individuals. Define the set of cohorts as A, where $a \subset A$ represents the cohort with age a. Let the set of individuals in this cohort be $N_a \subset N$, and let the number of individuals in such a set be defined as $|N_a|$. The probability a witnessed innovation is produced by an individual in the cohort with age a is,

$$\Pr(a) = \frac{\sum_{\{i \in N_a\}} x_i}{\sum_{\{i \in N\}} x_i} = \frac{|N_a|\, \bar{x}_a}{\sum_{\{a \in A\}} |N_a|\, \bar{x}_a}$$

where $\bar{x}_a$ is the average innovation potential of individuals in the cohort with age a. Dividing top and bottom by the size of the entire population, $|N|$, and defining the age distribution of the population as $p_a = |N_a| \,/\, |N|$, we rewrite this expression as,

$$\Pr(a) = \frac{p_a \bar{x}_a}{\sum_{\{a \in A\}} p_a \bar{x}_a}$$

which is the basis for (2) in the text.

Now consider the basis for the sub-model of cohort innovation potential, represented by (3) and repeated here

$$\bar{x}_a = L_1(a) L_2(a)$$

To derive this expression, assume that innovators start their lives with a period of education during which they do not innovate. Let the (stochastic) length of education for individual i be e_i. Additionally, define $g(a_i; z_i)$ as the individual's innovation potential if fully educated, where z_i is some (stochastic) measure of talent, effort, health, and any other factor that influences innovative ability. The innovation potential of individual i as a function of their age is then,

$$x_i = I(a_i \geq e_i) g(a_i; z_i)$$

where $I(a_i \geq e_i)$ is an indicator function equal to 1 if $a_i \geq e_i$ and 0 otherwise. Now employ a law of large numbers to write the cohort average innovation potential as,

$$\bar{x}_a(t) \xrightarrow{p} E[I(a_i \geq e_i) g(a_i; z_i)]$$

[31] One may think of innovators as being drawn, with replacement, from a box of names. A particular person's innovation potential then represents the frequency with which his or her name appears in the box, where we imagine that innovators with higher ability or effort level appear more often.

Assuming additionally that e_i and z_i are independent this expectation simplifies to,

$$\bar{x}_a(t) \xrightarrow{p} \Pr(a_i \geq e_i) E[g(a_i; z_i)] \tag{12}$$

giving the structure in (3). To estimate $L_1(a) = \Pr(a_i \geq e_i)$, we assume that e_i is distributed logistically within cohorts, and we similarly take a logistic specification for $L_2(a) = E[g(a_i; z_i)]$.

8 Data Appendix

This appendix describes the data sources used in the paper, providing both reference material and some underlying details of the methodology used in data collection.

8.1 Data on great innovators

A wealth of biographical information is available for Nobel Prize winners. The most useful immediate source is the official website of the Nobel Foundation, nobelprize.org. This website provides lists and written biographies of all winners, and was used to obtain dates and locations of birth, the field of the prize, the year and location of the highest educational degree, and the year(s) in which the prize-winning research was performed. Altogether, I was able to determine dates of birth for all 547 Nobel Prize winners between 1901 and 2003, and the period of key research for all but 3 of these. In practice, the data identifies a single year of great achievement - i.e. the year of success - for 75% of the cases. In the remainder of cases, the achievement for which the Nobel is awarded appears to encompass multiple sub-contributions, in which cases early and late dates of achievement were collected. In these cases, the estimations in the text use the middle year to define the age at great achievement, although results using either the first or last year of the key research are extremely similar in general and, in particular, nearly identical in the size of the trends and their statistical significance. The year or period of key research was usually straightforward to ascertain through the Nobel Foundations biographies, but in cases where these did not accurately identify the year or period of key research, other sources were consulted. The primary printed materials used were:

Schlessinger, B. and Schlessinger, J. *The Who's Who of Nobel Prize Winners*, 1901-1995. Oryx Press, Phoenix AZ 1996.

which was cross-referenced with,

Daintith, J. and Gjertsen, D. *The Grolier Library of Science Biographies.* Vols. 1-10. Grolier Educational, Danbury CT 1996.

Debus, A.G. ed. *World Who's Who in Science: A Biographical Dictionary of Notable Scientists from Antiquity to the Present.* Marquis Who's Who Inc., Chicago 1968.

McMurray, E.J., Kosek, J.K., and Valade, R.M. *Notable Twentieth-Century Scientists.* Vols. 1-4. Gale Research, Detroit 1995.

Williams, T.I. ed. *Biographical Dictionary of Scientists.* John Wiley and Sons, New York 1974.

Data on great inventors were collected from two technological almanacs that provide, by year, a list of notable technological advances in that year. These almanacs typically provide the date and location of birth of the innovator responsible, providing a set of 286 inventors in the 20th Century. The almanacs used were,

Bunch, B. and Hellemans, A. *The Timetables of Technology*. Simon and Schuster, New York 1993.

Ochoa, G. and Corey, M. *The Timeline Book of Science*. Ballantine Books, New York 1995.

The field of research is given according to set categories in both the Bunch et al and Ochoa et al sources. I condense the categorizations across these sources into nine fields: Communication, Computers & Electronics, Energy, Food & Agriculture, Materials, Medicine, Tools & Devices, Transportation, and Other. These categorizations define the field fixed effects in the econometric specification (1), but the results are not sensitive to specific categorizations.

8.2 Data on great athletes

For the Most Valuable Player Award in baseball, dates of birth and dates of the award were taken electronically from the *Lahman Baseball Database*, version 5.1, which was released on January 24, 2004. This database contains an enormous set of data for baseball players from 1871 through the present. It can be found at http:/baseball1.com/.

For world record breakers in track & field, dates of birth, dates of record breaking event, and nationalities were taking from the following source:

Lawson, G. *World Record Breakers in Track and Field Athletics*, Human Kinetics, Champaign, IL 1997.

8.3 Data on population age distribution

One and five percent micro-samples of the U.S. census are available electronically through IPUMS, the Integrated Public Use Microdata Series, which is maintained by the University of Minnesota:

Steven Ruggles, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander. *Integrated Public Use Microdata Series:* Version 3.0 [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor], 2004.

The smallest sample used was for the 1900 census, whose micro-sample provided data on approximately 100,000 individuals. The largest sample used was for the 2000 census, whose micro-sample provided data on approximately 2.8 million individuals. Existing census research available on the website (www.ipums.umn.edu/usa/chapter3/chapter3.html) indicates that these micro-samples provide accurate estimates of the population at large with regard to age. Population data for years in between decennial census years were determined by linear interpolation. Data for the year 1930 are not available, requiring interpolation for years between 1920 and 1940.

The subgroup of active workers is defined by having active labor force participation (LABFORCE=1) in the IPUMS data. The IPUMS attempts to recode census data to allow comparisons over time on a common basis, even if the census questions asked are not entirely consistent. From 1940, LABFORCE=1 for any individual who is actively working or seeking work. In 1900, the variable requires that individuals report any profession and have worked in the last 12 months. In 1910-1920, it includes those who report any "gainful occupation". This subsample is still large, with a minimum of 40,000 observations in 1900 and 1.3 million observations in 2000.

The subgroup of professional scientists and engineers requires further construction. The IPUMS uses the variable OCC1950 to define occupations across census years according to a common set of categories. I then take a subsample of these occupation codes that include the following relevant descriptions: professors and instructors in all subjects except social sciences (codes 012-026); engineers (codes 041-049); and natural scientists (codes 061-069). These data have two potential difficulties: first, the samples are substantially smaller, with only 56 observations in 1900 and 353 in 1910, rising to approximately 25,000 in the year 2000; second, the occupation is not defined until it is begun, in which case those still in school are not included. To create reasonable population estimates I first smooth these population data with an Epanechnikov kernel and a bandwidth of 2 years. Second, I impute the number of innovators still in training (those aged 15-29) based on the number of employed workers ten years later who are ten years older. The results are not sensitive to particular kernel bandwidths or age imputation schemes. In results not reported, I have also considered a broader set of all "professional, technical" workers (codes 001-099), which gives similar results.

References

[1] Adams, James D., Black, Grant C., Clemmons, J.R., and Stephan, Paula E. "Scientific Teams and Institutional Collaborations: Evidence from U.S. Universities, 1981-1999", NBER Working Paper #10640, July 2004.

[2] Bloom, B. "Generalizations about Talent Development," in *Developing Talent in Young People*, New York: Ballentine Books, 1985.

[3] Deaton, Angus. *The Analysis of Household Surveys*, Baltimore, MD: The Johns Hopkins University Press, 1997.

[4] Ericsson, K.A. and Lehmann, A. C. "Expert and Exceptional Performance: Evidence of Maximal Adaptation to Task Constraints," *Annual Review of Psychology*, 1996, 47, 273-305.

[5] Evenson, Robert E. "Patent Data by Industry: Evidence for Invention Potential Exhaustion?" *Technology and Productivity: The Challenge for Economic Policy*, 1991, Paris: OECD, 233-248

[6] Fan, Jianqing. "Design-adaptive Nonparametric Regression," *Journal of the American Statistical Association*, 1992, 87, 998-1004.

[7] Galenson, David W. "A Portrait of the Artist as a Young or Old Innovator: Measuring the Careers of Modern Novelists," NBER Working Paper #10213, 2004a.

[8] ——. "A Portrait of the Artist as a Very Young or Very Old Innovator: Creativity at the Extremes of the Lifecycle," NBER Working Paper #10515, 2004b.

[9] Galison, Peter, Gordin, Michael, and Kaiser, David, eds., *The History of Modern Physical Science in the Twentieth Century (Quantum Histories, vol. 4)*, New York: Routledge, 2001.

[10] Galenson, David W., and Weinberg, Bruce A. "Creating Modern Art: The Changing Careers of Painters in France from Impressionism to Cubism," *American Economic Review*, September 2001, 91 (4), 1063-1071.

[11] Jammer, Max. *The Conceptual Development of Quantum Mechanics*, New York: McGraw-Hill, 1966.

[12] Jones, Benjamin F. "The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder?" NBER Working Paper #11360, 2005.

[13] Jones, Benjamin F. and Weinberg, Bruce. "Age and Scientific Creativity", 2006, mimeo, Northwestern University.

[14] Jones, Charles I. "R&D-Based Models of Economic Growth," *Journal of Political Economy*, 1995a, 103, 759-784.

[15] ——. "Time Series Tests of Endogenous Growth Models," *Quarterly Journal of Economics*, 1995b, 110, 495-525.

[16] Kortum, Samuel S. "Research, Patenting, and Technological Change," *Econometrica*, November 1997, 65, 1389-1419.

[17] Kuhn, Thomas S. *The Structure of Scientific Revolutions*, Chicago: University of Chicago Press, 1962.

[18] Lehman, Harvey C. *Age and Achievement*, Princeton, NJ: Princeton University Press,1953.

[19] Lundqvist, Stig, ed. *Nobel Lectures, Physics 1971-1980*, Singapore: World Scientific Publishing Co., 1992.

[20] Machlup, Fritz. *The Production and Distribution of Knowledge in the United States*, Princeton, NJ: Princeton University Press, 1962.

[21] National Research Council, *On Time to the Doctorate: A Study of the Lengthening Time to Completion for Doctorates in Science and Engineering*, Washington, DC: National Academy Press, 1990.

[22] National Science Foundation, *Industrial Research and Development Information System*, Table H-19, www.nsf.gov/sbe/srs/iris/start.cfm, 2005.

[23] Reese, H,W. Lee, L.J., Cohen, S. H., and Puckett, J.M. "Effects of Intellectual Variables, Age, and Gender on Divergent Thinking in Adulthood", *International Journal of Behavioral Development*, November 2001, 25 (6), 491-500.

[24] Simonton, Dean K. *Scientific Genius: A Psychology of Science*, Cambridge: Cambridge University Press, 1988.

[25] ——. "Age and Outstanding Achievement: What Do We Know After a Century of Research?" *Psychological Bulletin*, September 1988, 104, 251-267.

[26] ——. "Career Landmarks in Science: Individual Differences and Interdisciplinary Contrasts," *Developmental Psychology*, January 1991, 27, 119-130.

[27] ——. "Creativity," in *The Encyclopedia of Gerontology*, San Diego, CA: Academic Press,1996.

[28] Stephan, Paula and Levin, Sharon. "Age and the Nobel Prize Revisited", *Scientometrics*, November 1993, 28, 387-399.

[29] Terman, F.E. "A Brief History of Electrical Engineering Education", *Proceedings of the IEEE*, August 1998, 86 (8), 1792-1800.

[30] Tilghman, Shirley (chair) et al. *Trends in the Early Careers of Life Sciences*, Washington, DC: National Academy Press, 1998.

[31] Weitzman, Martin L. "Recombinant Growth," *Quarterly Journal of Economics*, May 1998, 113, 331-360.

[32] Zuckerman, Harriet and Merton, Robert. "Age, Aging, and Age Structure in Science," in Merton, Robert, *The Sociology of Science*, Chicago, IL: University of Chicago Press, 1973, 497-559.

Table 1: Age Trends among Great Innovators

	Dependent Variable: Age at Great Achievement					
	Nobel Prize Winners			Great Inventors		
	(1)	(2)	(3)	(4)	(5)	(6)
Year of Great Achievement (in 100's)	5.83*** (1.37)	6.34*** (1.36)	7.79*** (1.54)	4.86** (2.31)	6.60** (2.58)	8.18** (3.29)
Field Fixed Effects	No	Yes	Yes	No	Yes	Yes
Country of Birth Fixed Effects	No	No	Yes	No	No	Yes
Number of observations	544	544	544	286	286	248
Time span	1873-1998	1873-1998	1873-1998	1900-1991	1900-1991	1900-1988
Average age	38.6	38.6	38.6	39.0	39.0	38.9
R^2	0.032	0.068	0.189	0.016	0.098	0.173

Notes: Coefficient on year of great achievement gives age trend in years per century. Standard errors are given in parentheses. Field fixed effects for Nobel Prizes comprise four categories: Physics, Chemistry, Medicine, and Economics. Field fixed effects for great inventors comprise nine categories: Communication, Electronics and Computers, Energy, Food and Agriculture, Materials, Medicine, Tools and Devices, Transportation, and Other.

** Indicates significance at a 95% confidence level.

*** Indicates significance at a 99% confidence level.

Table 2: Maximum Likelihood Estimation of Life-Cycle Innovation Potential

		(1)	(2)	(3)	(4)	(5)	(6)
Early Life Cycle Logistic Curve	μ_0 Initial Mean, in Years	24.0 (2.05)	23.3 (2.30)	23.4 (1.95)	22.6 (2.84)	22.8 (2.84)	25.8 (1.07)
	μ_1 Trend, in Years/Century	7.76 (3.22) [.016]	8.29 (3.49) [.018]	8.32 (2.71) [.002]	8.91 (5.23) [.088]	10.29 (4.56) [.024]	5.32 (1.97) [.007]
	ω Variance Parameter	2.40 (0.37)	2.47 (0.43)	2.38 (0.40)	2.43 (0.73)	2.30 (0.46)	2.21 (0.20)
Later Life Cycle Logistic Curve	θ_0 Initial Mean, in Years	45.5 (2.36)	46.6 (2.21)	50.0 (3.90)	53.7 (8.16)	43.7 (3.54)	46.9 (3.70)
	θ_1 Trend, in Years/Century	-0.00e-03 (8.63e-03)	-0.00e-03 (6.70e-03)	0.14e-03 (6.88e-03)	-4.61 (12.1)	0.00e-03 (1.28e-03)	-0.71 (5.17)
	ρ Variance Parameter	7.05 (1.10)	7.54 (0.91)	9.12 (1.60)	6.64 (2.72)	6.24 (0.99)	7.38 (7.40)
Data	Population	U.S. Population	Active Workers	Scientists and Engineers	U.S. Population	U.S. Population	U.S. Population
	Inventor Type	All	All	All	Technology Almanacs	Nobel Prize	All
	Nationality	US Born	US Born	US Born	US Born	US Born	All
	Number of Invention Observations	294	294	294	127	181	738
Log Likelihood		-1050.9	-1053.0	-1056.7	-463.6	-633.6	-2641.2

Notes: All estimates are maximum likelihood. Standard errors are given in parentheses and calculated using the inverse of the information matrix. P-values for the trend in the early life-cycle are given in square brackets.

Table 3: Maximum Likelihood Estimation: Further Specifications

		(1)	(2)	(3)	(4)
Early Life Cycle Logistic Curve	μ_0 Initial Mean, in Years	24.8 (1.94)	24.6 (1.84)	24.5 (2.36)	25.5 (1.19)
	μ_1 Trend, in Years/Century	6.32 (3.27) [.053]	6.05 (2.94) [.039]	6.47 (3.95) [.101]	5.89 (2.12) [.005]
	ω_0 Initial Variance Parameter	2.86 (0.83)	3.42 (1.01)	3.08 (1.14)	2.02 (4.18)
	ω_1 Trend, Variance Years/Century	-0.88 (1.42)	-1.74 (1.60)	-1.29 (1.80)	0.43 (7.35)
Later Life Cycle Logistic Curve	θ_0 Initial Mean, in Years	45.0 (4.49)	45.3 (3.22)	48.5 (4.88)	46.8 (3.79)
	θ_1 Trend, in Years/Century	0.99 (4.06)	2.29 (3.28)	2.79 (6.79)	-0.93 (2.53)
	ρ_0 Initial Variance Parameter	6.81 (2.06)	6.97 (1.47)	8.42 (2.13)	7.67 (0.73)
	ρ_1 Trend, Variance Years/Century	0.45 (1.77)	1.04 (1.46)	1.27 (3.07)	-0.42 (1.09)
Data	Population	Entire U.S. Population	All Active Workers	Scientists and Engineers	Entire U.S. Population
	Nationality	U.S. Born	U.S. Born	U.S. Born	All
	Number of Invention Observations	294	294	294	738
Log Likelihood		-1050.8	-1052.6	-1056.4	-2641.1

Notes: All estimates are maximum likelihood. Standard errors are given in parentheses and calculated using the inverse of the information matrix. P-values for the trend in the early life-cycle are given in square brackets.

Table 4: Age Trends at Highest Degree among Nobel Prize Winners

	Dependent Variable: Age at Highest Degree				
	(1)	(2)	(3)	(4)	(5)
Year of Highest Degree (in 100's)	4.11*** (0.61)	3.85*** (0.62)	3.86*** (0.62)	4.39*** (0.65)	3.22*** (1.22)
Field Fixed Effects	No	No	Yes	Yes	Yes
Country of Degree Fixed Effects	No	No	No	Yes	--
Data	All	Doctorate Only	All	All	U.S. Degree
Number of observations	505	484	505	505	213
Time span	1858-1990	1858-1990	1858-1990	1858-1990	1888-1990
Average age	26.5	26.6	26.5	26.5	26.6
R^2	.084	.075	.096	0.283	.060

Notes: Coefficient on year of highest degree gives age trend in years per century. Standard errors are given in parentheses. Field fixed effects for Nobel Prizes comprise four categories: Physics, Chemistry, Medicine, and Economics.

** Indicates significance at a 95% confidence level.

*** Indicates significance at a 99% confidence level.

Table 5: War Interruptions

	Ph.D Age		Lag between Degrees		Achievement Age	
	(1)	(2)	(3)	(4)	(5)	(6)
WW2	1.941*** (0.558)	2.034*** (0.522)	2.835*** (0.544)	2.797*** (0.517)	2.286** (1.148)	2.727** (1.187)
WW1	2.339** (1.149)	1.911* (1.104)	2.790** (1.125)	2.194** (1.096)	0.374 (2.723)	0.335 (2.830)
Year of Doctorate	0.037*** (0.006)	0.042*** (0.006)	0.021*** (0.008)	0.027*** (0.008)	--	--
Year of Great Achievement	--	--	--	--	0.068*** (0.014)	0.079*** (0.016)
Missing Educational Observation	-0.634* (0.366)	-1.098*** (0.403)	--	--	2.111** (0.834)	0.452 (0.985)
Field Fixed Effects	No	Yes	No	Yes	No	Yes
Country of Birth Fixed Effects	No	Yes	No	Yes	No	Yes
Number of observations	508	508	348	348	544	544
R^2	0.12	0.37	0.10	0.36	0.05	0.20

	Probability of Invention Between ages 25 and 30			
	(7)	(8)	(9)	(10)
Treatment	-.098*** (.024)	-.108*** (0.019)	-0.107** (0.023)	-0.110*** (0.020)
Control	.060 (.048)	.160*** (.060)	0.070 (0.056)	0.176*** (0.068)
Year of Great Achievement	-.0013*** (.0004)	-.0015*** (.0004)	-.0015*** (.0005)	-.0017*** (.0005)
Sample Mean of Dependent Variable	0.117	0.117	0.128	0.128
War	WW1	WW2	WW1	WW2
Field Fixed Effects	No	No	Yes	Yes
Country of Birth Fixed Effects	No	No	Yes	Yes
Number of observations	780	780	664	664
Log Likelihood	-274.1	-270.6	-237.1	-234.4

Notes, Left Panel: Results are OLS with standard errors in parentheses. WW2 and WW1 are dummies equal to 1 for individuals who happened to be between their undergraduate and graduate degrees at the outset of the indicated war (1939 for WW2 and 1914 for WW1). *Notes, Right Panel*: Results are for Probit model with coefficients reporting marginal probabilities and standard errors in parentheses. Treatment is a dummy equal to 1 for individuals born 1893-1898 (World War 1 specifications) and 1920-1925 (World War 2 specifications). Control is a dummy equal to 1 for individuals born 1893-1903 (World War 1) and 1920-1930 (World War 2). Field fixed effects comprise five categories: Physics, Chemistry, Medicine, and Economics for Nobel Prizes and a separate category for great inventors. Specifications (9) and (10) have fewer observations because some country of birth fixed effects predict failure perfectly and these observations are dropped. * Indicates significance at a 90% confidence level. ** Indicates significance at a 95% confidence level. *** Indicates significance at a 99% confidence level.

Figure 1: The Age Distribution of Great Innovation

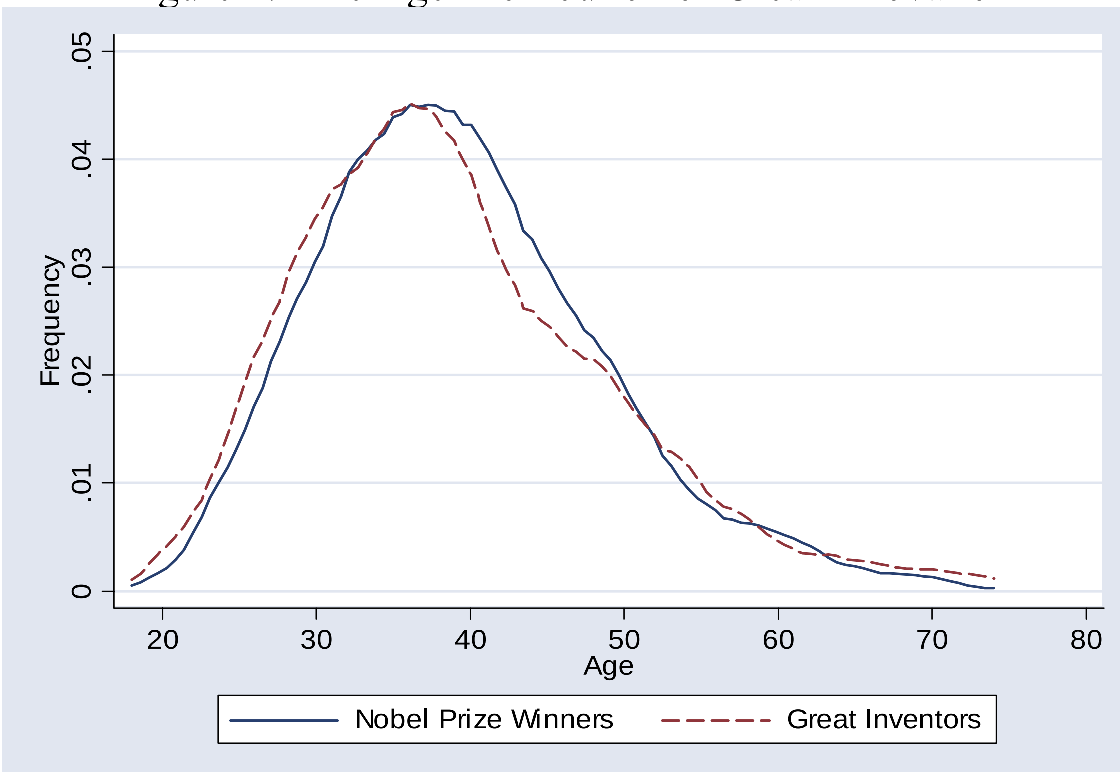

Figure 2: Shifts in the Age Distribution of Great Innovation

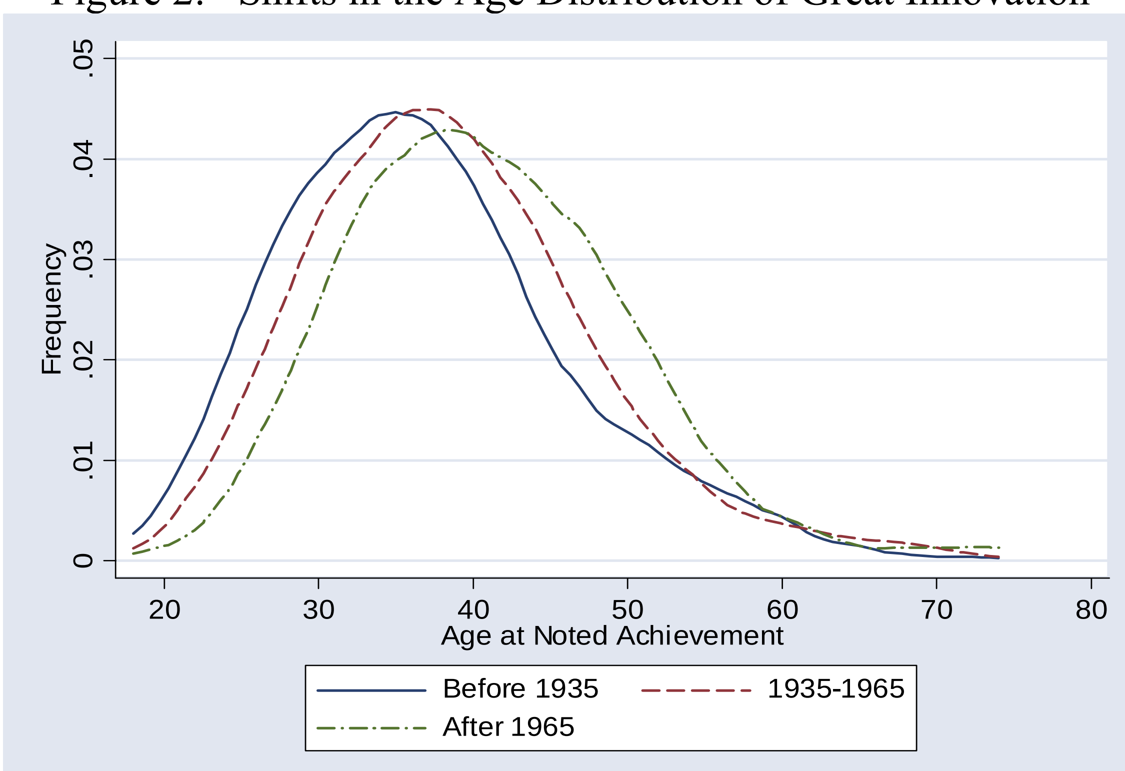

Figure 3: The Age Distribution of Baseball's Most Valuable Player

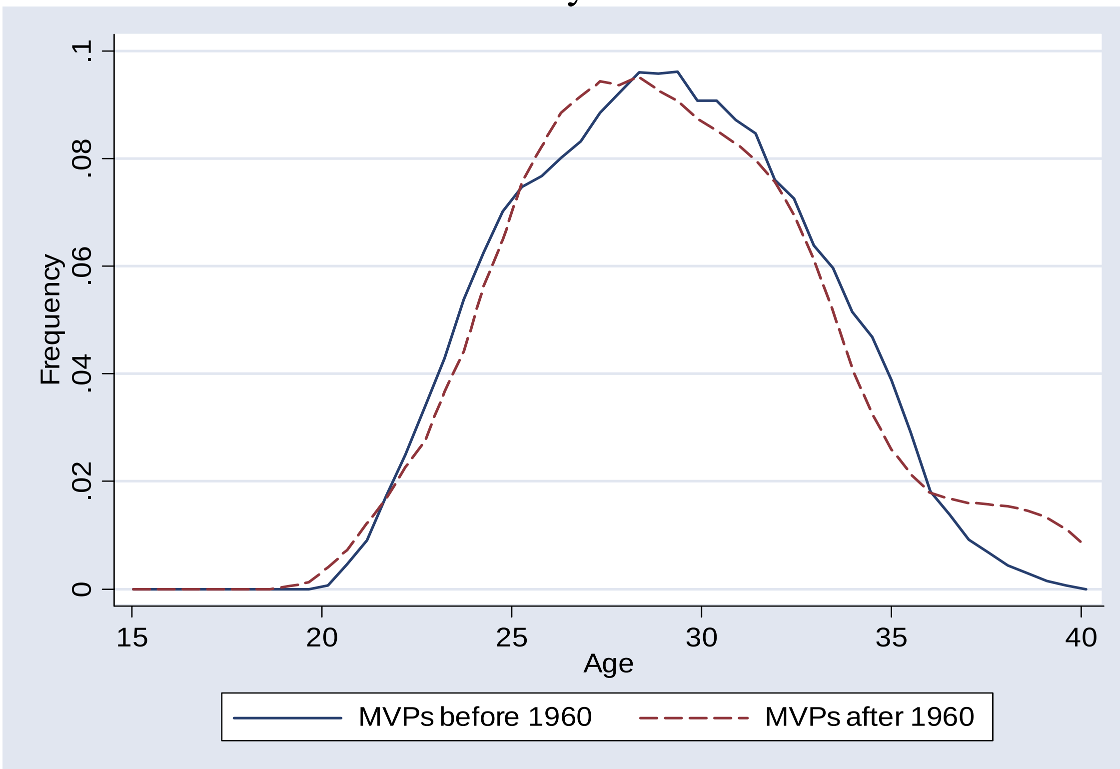

Figure 4: Model of Innovation Potential over the Life-Cycle

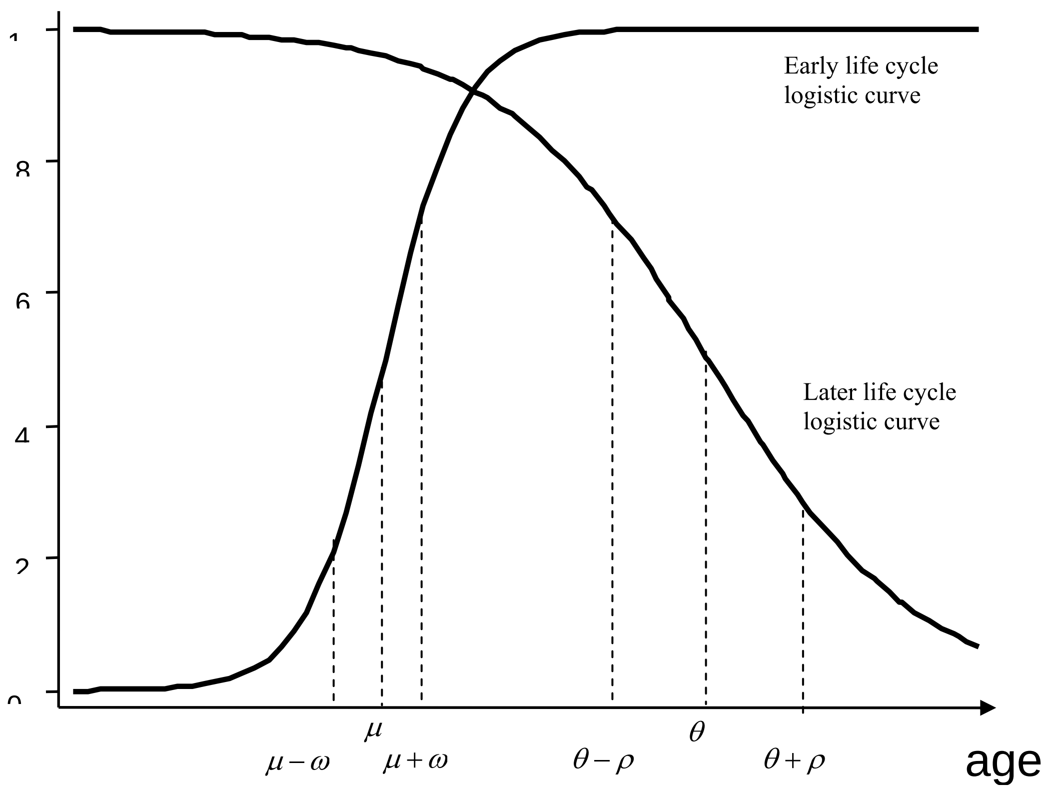

Figure 5: Maximum Likelihood Estimates for the Potential to Produce Great Innovations as a Function of Age

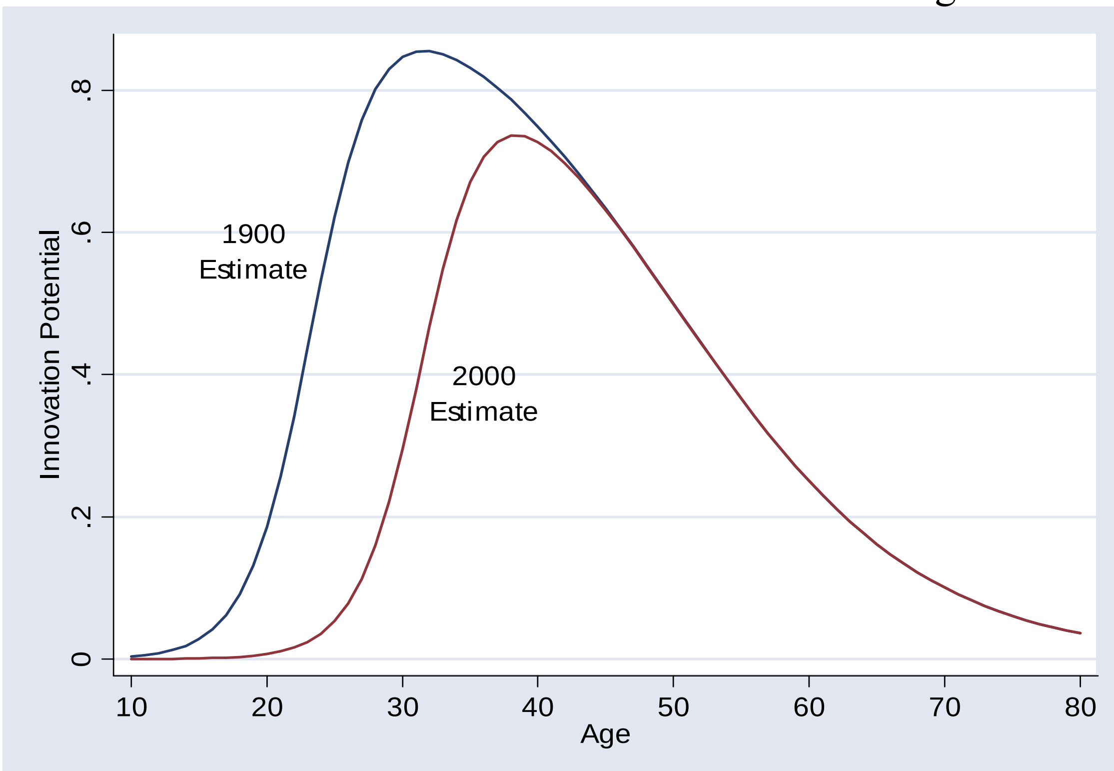

Figure 6: World War Interruptions and the Decline of Young Innovators

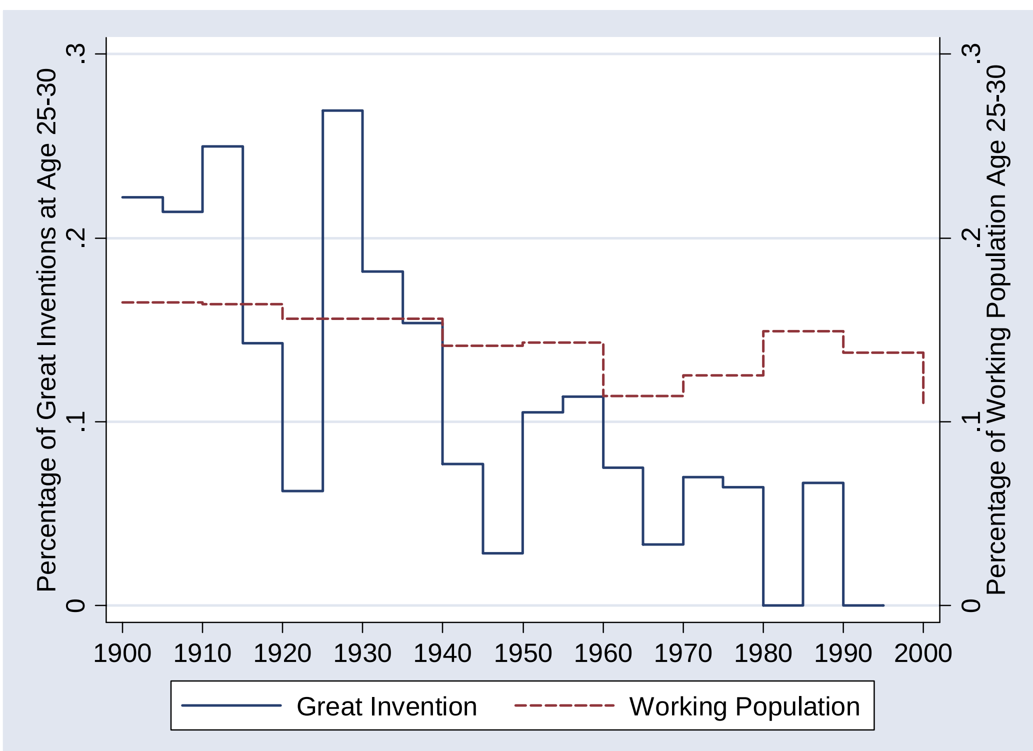

Figure 7: Ages at Ph.D. and Achievement over Time, by Field

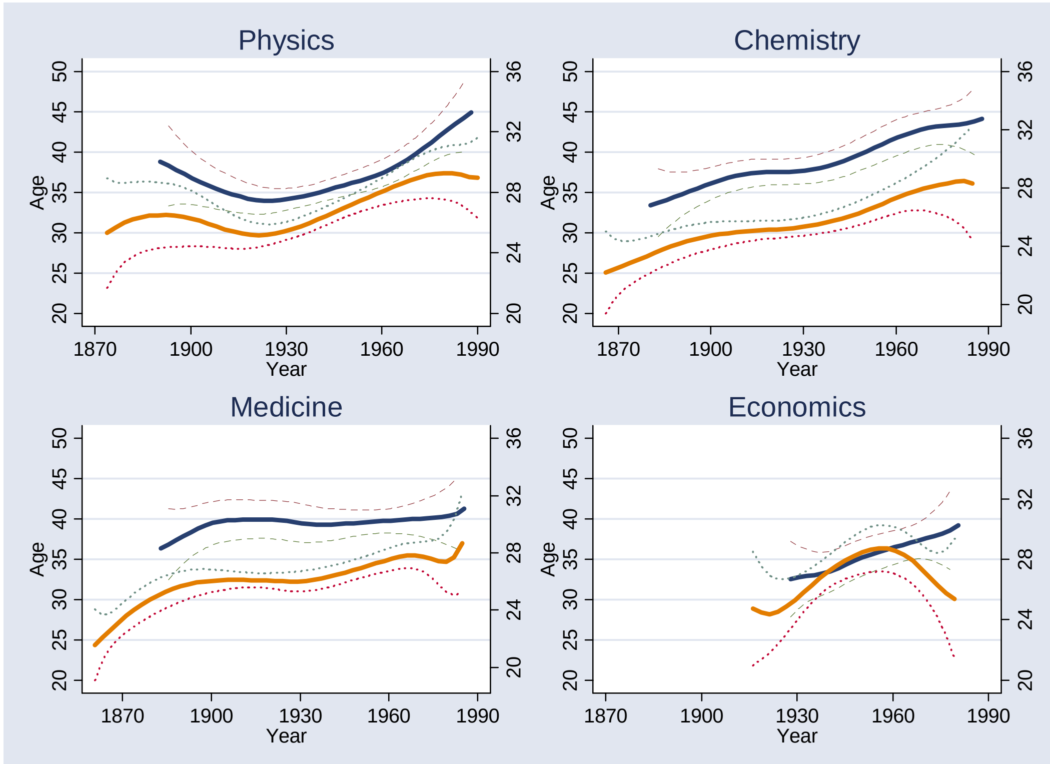

Figure 8: The Equilibrium Choice of Educational Attainment