THE REVIEW OF ECONOMIC STUDIES

Review of Economic Studies (2009) **76**, 283–317 0034-6527/09/00100283$02.00

The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder?

BENJAMIN F. JONES
Northwestern University and NBER

First version received May 2005; *final version accepted June* 2008 (*Eds.*)

This paper investigates a possibly fundamental aspect of technological progress. If knowledge accumulates as technology advances, then successive generations of innovators may face an increasing educational burden. Innovators can compensate through lengthening educational phases and narrowing expertise, but these responses come at the cost of reducing individual innovative capacities, with implications for the organization of innovative activity—a greater reliance on teamwork—and negative implications for growth. Building on this "burden of knowledge" mechanism, this paper first presents six facts about innovator behaviour. I show that age at first invention, specialization, and teamwork increase over time in a large micro-data set of inventors. Furthermore, in cross-section, specialization and teamwork appear greater in deeper areas of knowledge, while, surprisingly, age at first invention shows little variation across fields. A model then demonstrates how these facts can emerge in tandem. The theory further develops explicit implications for economic growth, providing an explanation for why productivity growth rates did not accelerate through the 20th century despite an enormous expansion in collective research effort. Upward trends in academic collaboration and lengthening doctorates, which have been noted in other research, can also be explained in this framework. The knowledge burden mechanism suggests that the nature of innovation is changing, with negative implications for long-run economic growth.

1. INTRODUCTION

Understanding innovation is central to understanding many important aspects of economics, from market structure to aggregate growth. Innovators, in turn, are a necessary input to any innovation. The innovator, wrestling with a creative idea, working with colleagues, and bringing an idea to fruition, seems the very heart of the innovative process.

This paper places innovators at the centre of analysis and focuses on two simple observations. First, innovators are not born at the frontier of knowledge; rather, they must initially undertake significant education. Second, the frontier of knowledge varies across fields and over time. This paper presents facts and theory that build on these observations, suggesting possibly fundamental consequences for the organization of innovative activity and, in the aggregate, for growth.

The first observation concerns human capital and highlights a general distinction between human capital and other stock variables. Physical stocks can be transferred easily, as property rights, from one agent to another. Human capital, by contrast, is not transferred easily. The vessel of human capital—the individual—is born with little knowledge and absorbs information at a limited rate, so that training occupies a significant portion of the life cycle. The difficulty of transferring human capital has broad implications in economics;[1] in this paper, I focus on basic implications for innovation.

1. See, for example Ben-Porath (1967) regarding life cycle earnings and Hart and Moore (1994) regarding debt contracts.

The second observation concerns the total stock of knowledge. In 1676, Isaac Newton wrote famously to Robert Hooke, "If I have seen further it is by standing on ye sholders of Giants". Newton's sentiment suggests that knowledge begets new knowledge, an observation that has been formalized in the growth literature (Romer, 1990; Jones, 1995*a*; Weitzman, 1998), with implications discussed extensively both there (*e.g.* Jones, 1995*b*; Kortum, 1997; Young, 1998) and in the micro-innovation literature (*e.g.* Scotchmer, 1991; Henderson and Cockburn, 1996). This paper suggests a different, indirect implication of Newton's observation: if one is to stand on the shoulders of giants, one must first climb up their backs, and the greater the body of knowledge, the harder this climb becomes.

If innovation increases the stock of knowledge, then the educational burden on successive cohorts of innovators may increase. Innovators might confront this difficulty through two basic margins. First, they may choose to learn more. Second, they might compensate by choosing narrower expertise. Choosing to learn more will leave less time in the life cycle for innovation. Narrowing expertise, meanwhile, can reduce individual capabilities and force innovators to work in teams. Intriguing evidence along the lines of a "learning more" effect is shown in Table 1, which borrows from Jones (2005*a*) and documents a rising age at great achievement and rising doctoral age among Nobel Prize winners over the 20th century. To help motivate the specialization margin, and the resulting need for teamwork, consider the invention of the micro-processor. As described by Malone (1995), the invention was by necessity the work of a team. The inspiration began with a researcher named Ted Hoff, who joined in the development with Stan Mazor. But as Malone writes,

> Hoff and Mazor didn't really know how to translate this architecture into a working chip design .... In fact, probably only one person in the world did know how to do the next step. That was Federico Faggin ....

The micro-processor was one person's inspiration but several people's invention. It is the story of researchers with circumscribed abilities, working in a team, and it helps motivate the investigations of this paper.

I begin below by presenting six facts. Using a rich patent data set (Hall, Jaffe and Trajtenberg, 2001) together with the results of a new data collection exercise to determine the ages of 55,000 inventors, I develop detailed patent histories for individuals. I show that (i) the age at first invention, serving as a proxy measure for educational attainment; (ii) a measure of specialization; and

TABLE 1

Age trends among Nobel Prize winners

	Dependent variable	
	Age at great achievement	Age at highest degree
Year of great achievement	5·83***	—
(in 100's)	(1·37)	
Year of highest degree	—	4·11***
(in 100's)		(0·61)
Number of observations	544	505
Time span	1873–1998	1858–1990
Average age	38·6	26·5
R^2	0·027	0·084

Notes: (i) This table borrows from Jones (2005*a*). Age trends are measured in years per century. S.E. are given in parentheses. (ii) Nobel Prize winners include all winners in Physics, Chemistry, Medicine, and Economics. Age at great achievement is age when contribution is made (not later age when prize is awarded).
***Indicates 99% confidence level.

(iii) team size are all increasing over time at substantial rates (Figure 1). These trends are robust to a number of controls and in particular are robust across a wide range of technological categories and research environments. An informal theory of the "burden of knowledge" might suggest these effects. Innovators, when faced with greater knowledge depth, might respond through both longer educational periods and greater specialization.

In cross-section, I develop a measure of "knowledge depth" and show that (iv) teamwork and (v) specialization are greater in fields with deeper knowledge. Like the time series results, these cross-sectional patterns are robust to numerous controls and, furthermore, seem natural within an informal theory of the burden of knowledge. The final fact is then particularly surprising: (vi) the average age at first invention is strikingly similar across fields and does not vary with the depth of knowledge. This fact suggests a more nuanced mechanism, and the balance of the paper presents a model that ties these six facts together. I show how these facts can emerge in tandem, clarifying the influence of burden of knowledge on innovator behaviour and building precise implications for innovators' aggregate output and thus economic growth.

In the model, innovators are specialists who interact with each other in the implementation of their ideas. The model introduces different areas of application (*e.g.* airplanes or drugs) within which innovators define their specialties. Achieving expertise requires an innovator to bring himself or herself to the frontier of knowledge within some area of application, and the difficulty of reaching the frontier—the burden of knowledge—may vary across areas and over time.

The central choice problem is that of career. At birth, each individual chooses to become either a production worker or an innovator. Innovators must further choose specific knowledge

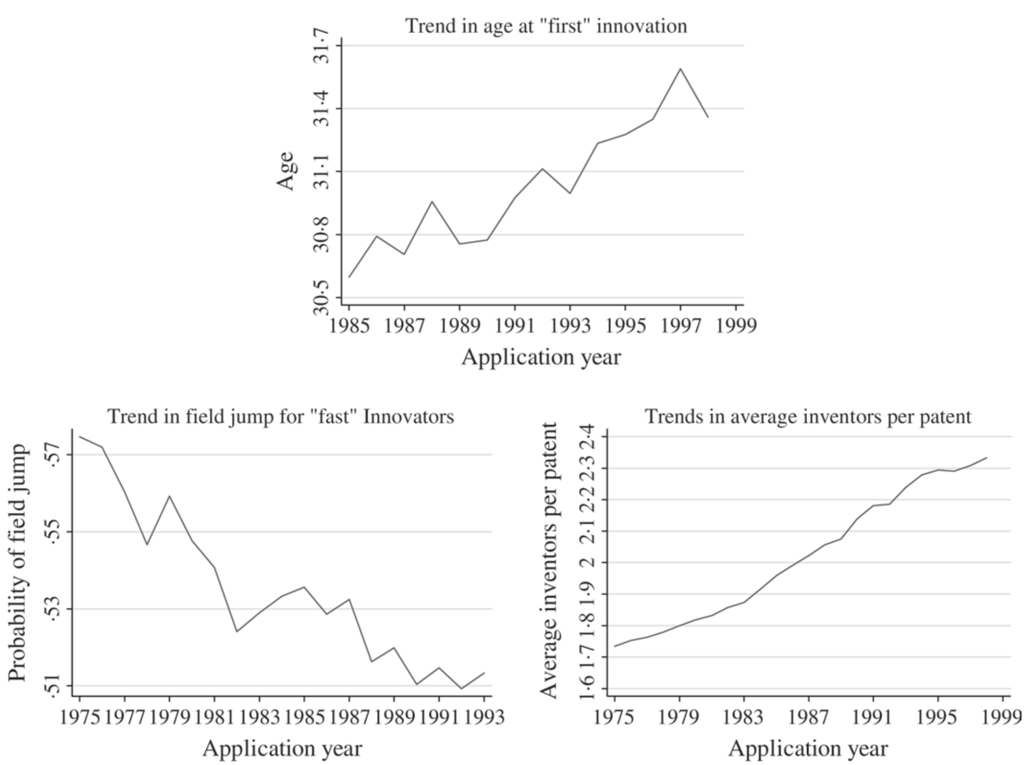

FIGURE 1

Basic time trends

to learn. This choice is partly one of specialization, with the innovator trading off the costs and benefits of broader education: more knowledge leads to increased innovative potential but also costs more to acquire. Crucially, however, the career choice is also one of application—what broad area of knowledge to enter (*e.g.* airplanes or drugs). In making this decision, innovators are attracted to areas with relatively low knowledge requirements and/or better opportunities, but they also seek to avoid crowding. Other things equal, the greater the duplication of innovators in a particular area of knowledge, the less expected income each will earn. This decision helps pin down innovator behaviour. In particular, arbitraging across different application areas, innovators allocate themselves to equate expected income across areas of research. Once income has been equalized, innovators find equivalent value in education and are only willing to undertake the same total education across widely different fields. It then falls to specialization to confront variation in the difficulty of reaching the knowledge frontier. Hence, the model predicts equivalent educational attainment in cross-section but increased specialization and teamwork in deeper areas of knowledge, as the facts suggest.

The time series behaviours and growth implications emerge in the dynamic features of the economy. The model marries the burden of knowledge mechanism to two other dimensions—population growth and technological opportunity—that are much discussed in the existing growth literature. First, a growing population allows the economy to continuously scale up innovative effort, keeping growth going even as individual contributions are in decline (as seen in Jones, 1995*a*). In this model, population growth also plays a key role by increasing the market size for innovations and thus the marginal benefit of education. Second, technological opportunities may rise or fall as the economy evolves. This feature captures, in reduced form, a broad range of arguments in the literature: both "fishing-out" arguments (*e.g.* Kortum, 1997) and more optimistic specifications where innovation is increasingly easy (*e.g.* Romer, 1990; Aghion and Howitt, 1992). In the model, changing technological opportunities, like population growth, also affect the marginal benefit of education.

In this framework, the same forces that influence innovators' educational decisions also influence long-run growth. Indeed, individuals' educational decisions are made in the context of shifting knowledge burden, market size, and technological opportunities, producing detailed predictions about innovator behaviour on the one hand and aggregate consequences on the other hand. I show that, along a balanced growth path, innovators will seek more education with time, with increasing specialization and teamwork driven by a rising burden of knowledge. The model can thus explain the time series patterns of innovator behaviour (Figure 1). Moreover, the balanced growth path is explicitly determined, with the burden of knowledge seen to act on growth similarly to the fishing-out effect of more standard models. Therefore, one may view the burden of knowledge as a micro-foundation for fishing-out-type effects on growth. Alternatively, if one is convinced that a fishing-out process operates independently, then the burden of knowledge is seen as an additional effect constraining the growth rate.

The model can thus serve as a parsimonious explanation for the six facts about the micro-behaviour of innovators identified in this paper. As discussed in Section 4, the model can further explain several facts documented elsewhere, including upward trends in academic co-authorship and doctoral duration. Last, the model provides one consistent explanation for important aggregate data patterns. First, R&D employment in leading economies has been rising dramatically; yet, Total Factor Productivity (TFP) growth has been flat (Jones, 1995*b*). Second, the average number of patents produced per R&D worker or R&D dollar has been falling over time across countries (Evenson, 1984) and U.S. manufacturing industries (Kortum, 1993). This absence of "scale effects" in growth is much debated in the growth literature. It can be understood through the model as a burden of knowledge effect, building growth on foundations that also support a consistent interpretation for the micro-evidence presented in this paper.

This paper is organized as follows. Section 2 presents six central facts about the behaviour of innovators. Section 3 presents the burden of knowledge model, which ties these facts together and considers the growth implications. Section 4 discusses further empirical applications and generalizations of the theory, together with concluding comments.

2. ECONOMETRIC EVIDENCE

This section presents a set of facts about the behaviour of innovators. Using an augmented patent data set, we will be able to examine three outcomes in particular:

1. Team size
2. Age at first innovation
3. Specialization.

The data are described in the following subsection. An investigation of basic time trends and cross-sectional results follows.

2.1. *Data*

I make extensive use of a patent data set put together by Hall *et al.* (2001). This data set contains every utility patent issued by the U.S. Patent and Trademark Office (USPTO) between 1963 and 1999. The available information for each patent includes (i) the grant date and application year and (ii) the technological category. The technological category is provided at various levels of abstraction: a 414 main patent class definition used by the USPTO as well as more organized 36-category and 6-category measures created by Hall *et al.* (The 36-category and 6-category measures are described in Table 5.) For patents granted after 1975, the data set includes additionally (iii) every patent citation made by each patent and (iv) the names and addresses of the inventors listed with each patent. There are 2·9 million patents in the entire data set, with 2·1 million patents in the 1975–1999 period (Figure 2).

Using the data available over the 1975–1999 time period, we can define two useful measures directly:

- Team size: The number of inventors listed with each patent.
- Time lag: The delay between consecutive patent applications from the same inventor.

For the latter measure, we identify inventors by their last name, first name, and middle initial, and then build detailed patent histories for each individual.

We can also define two more approximate measures that will be useful for analysis:

- Tree size: The size of the citations "tree" behind any patent. Any given patent will cite a number of other patents, which will in turn cite further patents, and so on. For the purposes of cross-sectional analysis, the number of nodes in a patent's backwards-looking patent tree serves as a proxy measure for the amount of underlying knowledge.

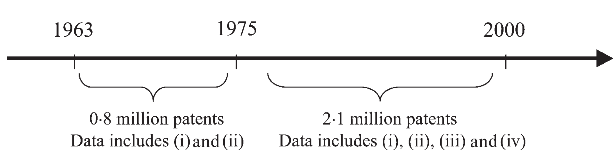

FIGURE 2

Summary of available data

- Field jump: The probability that an innovator switches technological areas between consecutive patent applications. This can serve as a proxy measure for the specialization of innovators. The more specialized you are, the less capable you are of switching fields.

A limitation of this last measure is that since technological categories are assigned to patents and not to innovators, inferring an innovator's specific field of expertise is difficult when innovators work in teams. For inventors who work in teams, the relation between specialization and field jump is in fact ambiguous: as inventors become more specialized and work in larger teams, they may jump as regularly as they did before. For the specialization analysis, we therefore focus on solo inventors, for whom increased specialization is associated with a decreased capability of switching fields.

Finally, we would like to investigate the age at first innovation, as an outcome-based measure that delineates the pre-innovation and active innovation phases. Unfortunately, inventors' dates of birth are not available in the data set, nor from the USPTO generally. However, using name and zip code information, it was possible to attain birth date information for a large subset of inventors through a public Web site, http://www.AnyBirthday.com. The Web site http://www.AnyBirthday.com uses public records and contains birth date information for 135 million Americans. The Web site requires a name and zip code to produce a match. Using a Java program to repeatedly query the Web site, it was found that of the 224,152 inventors for whom the patent data included a zip code, http://www.AnyBirthday.com produced a unique match in 56,281 cases. The age data subset and associated selection issues are discussed in detail in Appendix B. The analysis given there shows that the age subset is not a random sample of the overall innovator population. This caveat should be kept in mind when examining the age results, although it is mitigated by the fact that the differences between the groups become small when explained by other observables, controlling for these observables in the age regressions has little effect, and the results for team size and specialization persist when examining the age subset. See the discussion in Appendix B.

2.2. *Time series results*

I consider the evolution over time of our three outcomes of interest. Figure 1 presents the basic data, while Tables 2 through 4 examine the time trends in more detail.

Consider team size first. The lower right panel of Figure 1 shows that team size is increasing at a rapid rate, rising from an average of 1·73 in 1975 to 2·33 at the end of the period, for a 35% increase overall. Table 2 explores this trend further by performing regressions relating team size to application year, and we see that the time trend is robust to a number of controls. Controlling for compositional effects shows that any trends into certain technological categories or towards patents from abroad have little effect. Repeating the regressions separately for patents from domestic versus foreign sources shows that the domestic trend is steeper, though team size is rising substantially regardless of source. Repeating the time trend regression individually for each of the 36 different technological categories defined by Hall *et al.* shows that the upward trend in team size is positive and highly significant in every single technological category. Running the regressions separately by "assignee code" to control for the type of institution that owns the patent rights shows that the upward trend also prevails in each of the seven ownership categories identified in the data, indicating that the trend is robust across corporate, government, and other research settings, both in the U.S. and abroad.[2] In short, we find an upward trend in team size that is both general and steep.

2. Table B.2 describes the ownership assignment categories.

TABLE 2

Trends in inventors per patent

	Dependent variable: inventors per patent						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Application year	0·0293	0·0261	0·0262	0·0251	0·0244	0·0306	0·0180
	(0·0001)	(0·0001)	(0·0001)	(0·0001)	(0·0002)	(0·0002)	(0·0003)
Foreign patent	—	0·444	0·416	0·141	0·146	U.S. only	Foreign only
		(0·002)	(0·002)	(0·004)	(0·004)		
Technological field controls							
Broad	—	Yes	—	—	—	—	—
Narrow	—	—	Yes	Yes	Yes	Yes	Yes
Assignee code	—	—	—	Yes	Yes	Yes	Yes
Number of observations	2,016,377	2,016,377	2,016,377	2,016,377	1,506,956	1,123,310	893,067
Period	1975–1999	1975–1999	1975–1999	1975–1999	1975–1996	1975–1999	1975–1999
Mean of dependent variable	2·03	2·03	2·03	2·03	1·97	1·82	2·29
Per decade trend as % of period mean	14·4	12·9	12·9	12·4	12·4	16·8	7·9
R^2	0·02	0·08	0·10	0·12	0·13	0·12	0·10

Notes: (i) Regressions are ordinary least squares with S.E. in parentheses. Specifications (1) through (4) consider the entire universe of patents applied for between 1975 and 1999. Specification (5) considers only patents that were granted within 3 years after application (see discussion in text). Specifications (6) and (7) present separate trends for domestic and foreign source patents. (ii) Foreign patent is a dummy variable to indicate whether the first inventor listed with the patent has an address outside the U.S. (iii) "Broad" technological controls include dummies for each of the six categories in the most aggregated technological classification of Hall *et al.* "Narrow" technological controls include dummies for each category of their 36-category classification. (iv) Upward trends persist when run separately for each technological field. Using the broad classification (six categories), the trends range from a low of 0·018 for "Other" to a high of 0·037 for "Chemical". Using the narrower classification scheme (36 categories), the trends range from a low of 0·007 for "Apparel &Textile" to 0·051 for "Organic Compounds". The smallest t-statistic for any of these trends is 7·76. (v) Assignee code controls are seven dummy variables that define who holds the rights to the patent. Most patent rights are held by U.S. or foreign corporations (80%), while a minority remain unassigned (17%) at the time the patent is issued. Table B.2 describes the assignee codes in further detail. Running the time trends separately for the individual assignee codes shows that the team size trends range from a low of 0·005 for the unassigned category to a high of 0·039 for U.S. non-government institutions. The lowest t-statistic for any of these trends is 5·38.

Next, consider the age at first innovation. Note that we define an innovator's "first" innovation as the first time they appear in the data set. Since we cannot witness individuals' patents before 1975, this definition is dubious for (i) older individuals and (ii) observations of first innovations that occur close to 1975. To deal with these two issues, I limit the analysis to those people who appear for the first time in the data set between the ages 25 and 35 and after 1985. The upper panel of Figure 1 plots the average age over time, where we see a strong upward trend. The basic time trend in Table 3 shows an average increase in age at a rate of 0·66 years per decade. Controlling for compositional biases due to shifts in technological fields or team size has no effect on the estimates. The results are also similar when examining different age windows.[3] Analysis of trends within technological categories shows that the upward trend in age is quite general. Smaller sample sizes tend to reduce significance when the data is finely cut, but an upward age trend is found in all six technology classes using six-category measure of Hall *et al.*, and in 29 of 36 categories when using their 36-category measure. The upward age trend also persists across all patent ownership classifications.

Note that the age at first invention serves as an outcome-based measure to delineate the education phase and phase in the life cycle. A possible contaminating factor is the duration it takes to produce an innovation (the age at first invention is the sum of age at completion of education plus the time lag until the first invention). However, in results reported elsewhere (Jones, 2005*b*), the time lags between an inventor's inventions are short, do not trend over time, and vary only modestly across fields. Thus, the age at first invention appears to track the end of the educational phase with little error. Some related evidence regarding doctoral duration is considered in Section 4.[4]

Now we turn to specialization. The specialization measure considers the probability that an innovator switches fields between consecutive innovations. Before examining the raw data, it is necessary to consider a truncation problem that may bias us towards finding increased specialization over time. The limited window of our observations (1975–1999) means that the maximum possible time lag between consecutive patents by an innovator is largest in 1975 and smallest in 1999. This introduces a downward bias over time in the lag between innovations. It is intuitive, and it turns out in the data, that people are more likely to jump fields the longer they go between innovations.[5] Mechanically shorter lags as we move closer to 1999 can therefore produce an apparent increase in specialization. To combat this problem, I make use of a conservative and transparent strategy. I restrict the analysis to a subset of the data that contains only consecutive innovations that were made within the same window of time. In particular, we examine only the consecutive innovations when the second application comes within 3 years of the first. Furthermore, we examine only the innovations that were granted within 3 years of the application.[6] This strategy eliminates the bias problem at the cost of limiting our data analysis to the 1975–1993

3. The table reports results for the 23–33 age window as well. In results not reported, I find that the trend is similar across subsets of these windows: ages 23–28, 25–30, 31–35, and so forth. Furthermore, there is no upward trend when examining age windows beginning at age 35.

4. Doctoral age is also an imperfect delineation between education and innovation phases because doctorates explicitly require innovative research that begins well before the awarding of the degree.

5. An interpretation consistent with the spirit of the burden of knowledge concept is that people need time to reeducate themselves when they jump fields; hence, a field jump is associated with a larger time lag.

6. Examining only the patents where the second application came within 3 years limits our analysis to those cases where the first application was made before 1997. However, a second issue is that patents are granted with a delay—2 years on average—and only patents that have been granted appear in the data. For a first patent applied for in 1996, it is therefore much more likely that we will witness a second patent applied for in 1997 than one applied for in 1999—introducing further downward bias in the data. To deal completely with the truncation problem, we therefore further limit ourselves to patents that were granted within 3 years of their application, which means that we only examine the period 1975–1993.

TABLE 3

Trends in age at first innovation

	Dependent variable: age at application						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Application year	0·0657	0·0666	0·0671	0·0671	0·0687	0·0530	0·0584
	(0·0095)	(0·0095)	(0·0095)	(0·0099)	(0·0097)	(0·0107)	(0·0109)
Technological field controls							
Broad	—	Yes	—	—	—	—	—
Narrow	—	—	Yes	Yes	Yes	—	Yes
Assignee code	—	—	—	Yes	Yes	—	Yes
Team size	—	—	—	—	−0·0630	—	−0·0348
					(0·0273)		(0·0306)
Number of observations	6541	6541	6541	6541	6541	5102	5102
Period	1985–1999	1985–1999	1985–1999	1985–1999	1985–1999	1985–1999	1985–1999
Age range	25–35	25–35	25–35	25–35	25–35	23–33	23–33
Mean of dependent variable	31·0	31·0	31·0	31·0	31·0	29·3	29·3
Per decade trend as % of period mean	2·1	2·1	2·2	2·2	2·2	1·8	2·0
R^2	0·007	0·010	0·020	0·020	0·021	0·005	0·018

Notes: (i) Regressions are ordinary least squares, with S.E. in parentheses. All regressions observe only at those innovators for whom we have age data and who appear for the first time in the data set in or after 1985. Specifications (1) through (5) consider those innovators who appear for the first time between ages 25 and 35. Specifications (6) and (7) consider those innovators who appear for the first time between ages 23 and 33. (ii) Broad technological controls include dummies for each of the six categories in the most aggregated technological classification Hall *et al.* Narrow technological controls include dummies for each classification in their 36-category measure. The upward age trend persists when run separately in each of broad technology classes of Hall *et al.* These trends are significant in five of the six categories, with similar trend coefficients as when the data are pooled. Upward trends are also found in 29 of the 36 categories when using the narrow technology classification of Hall *et al.* Here, 12 categories show significant upward trends. Sample sizes drop considerably when the data are divided into these 36 categories. The one case of a significant downward trend (category no. 23, Computer Peripherals) has 42 observations. (iii) Assignee code controls are seven dummy variables that define who holds the rights to the patent. Table B.2 describes the assignee codes in further detail. The upward age trends persist when run separately for each assignee code and are similar in magnitude to the trends in Table 2.

period and making our results applicable only to the subsample of "faster" innovators.[7] The lower left panel of Figure 1 shows the trend from 1975 to 1993.

Table 4 considers the trend in specialization with and without this corrective strategy. The results there, together with the graphical presentation in Figure 1, indicate a smooth decrease in the probability of switching fields. The decline is again quite steep. Using the central estimate for the trend of $-0{\cdot}003$, we can interpret a 6% increase in specialization every 10 years. Note that our main results, and Figure 1, use the 414-category measure for technology to determine whether a field switch has occurred. This is our most accurate measure of technological field (the measures of Hall *et al.* are aggregations of it), but the results are not influenced by the choice of field measure. Note in particular that the *percentage* trend is robust to the choice of the 6-category, 36-category, or 414-category measure for technology—the trend is approximately 6% per decade for all three. Including controls for U.S. patents, the application time lag, ownership status, and the technological class of the initial patent has little effect. Furthermore, examining for trends within each of the 36 categories of Hall *et al.*, we find that the probability of switching fields is declining in 34 of the 36; the decline is statistically significant in 20. In sum, we see a robust and strongly decreasing tendency for solo innovators to switch fields.

2.3. *Cross-section results*

For a first observation of the data in cross-section, Table 5 presents a simple comparison of means across the 6 and 36 technological categories of Hall *et al.* (2001). The middle column in the table presents the mean age at first innovation, and the data show a remarkable consistency across technological categories. In 31 of the 36 categories, an innovator's first innovation tends to come at age 29. The lowest mean age among the 36 categories is 28·8, and the highest—an outlier that relies on only 12 observations—is 31·1. The table shows that regardless of whether the invention comes in "Nuclear & X-rays", "Furniture, House Fixtures", "Organic Compounds", or "Information Storage", the mean age at first innovation is nearly the same.[8]

The next columns of the table consider the average team size. Here, we see large differences across technological areas. The largest average team size, 2·91 for the "Drugs" subcategory, is over twice that of the smallest, 1·41 for the "Amusement Devices" subcategory.

Finally, the last columns of the table consider the probability that a solo innovator will switch subcategories between innovations. Here, as with team size and unlike the age at first innovation, we see large differences across technological areas. This variation is again consistent with the predictions of the model. At the same time, this basic, cross-sectional variation in the probability of field jump is difficult to interpret: the probability of field jump will be tied to how broadly a technological category happens to be defined, which may vary to a large degree across categories.

I can go further using a direct measure of the quantity of knowledge underlying a patent. In particular, I can analyse in cross-section what an increase in the knowledge measure implies for our outcomes of interest.

7. These restrictions maintain a significant percentage of the original sample. For example, of the 111,832 people who applied successfully for patents in 1975, 81,955 of them received a second patent prior to 2000. Of these 81,955 people for whom we can witness a time lag between applications, 79·8% made their next application within 3 years. Of those, 88·5% were granted both patents within 3 years of application.

8. These results can also be considered in a regression format. Pooling cross-sections and using application year dummies to take care of trends, the results are extremely similar. One can also adjust the time at first innovation by subtracting category-specific estimates of the time lag to get a closer estimate of an individual's education. One can also observe different age windows. The result that ages are nearly identical across fields is highly robust.

TABLE 4

Trends in probability of field jump

	Dependent variable: probability of switching technological field							
	(1) 414	(2) 414	(3) 36	(4) 36	(5) 6	(6) 6	(7) 414	(8) 414
Application year	−3·4e − 3	−3·2e − 3	−2·5e − 3	−2·8e − 3	−1·9e − 3	−2·3e − 3	−5·1e − 3	−3·0e − 3
	(0·19e − 3)	(0·19e − 3)	(0·19e − 3)	(0·19e − 3)	(0·17e − 3)	(0·17e − 3)	(0·12e − 3)	(0·11e − 3)
Foreign patent	—	0·0076	—	−0·0041	—	0·0002	—	−0·0005
		(0·0039)		(0·0038)		(0·0035)		(0·0029)
Time between applications	—	0·0225	—	0·0206	—	0·0154	—	0·0228
		(0·0012)		(0·0012)		(0·0011)		(0·0004)
Technological field controls (first patent)	—	Yes	—	Yes	—	Yes	—	Yes
Assignee code (first patent)	—	Yes	—	Yes	—	Yes	—	Yes
Number of observations	215,855	215,855	215,855	215,855	215,855	215,855	359,405	359,405
Period	1975–1993	1975–1993	1975–1993	1975–1993	1975–1993	1975–1993	1975–1999	1975–1999
Mean of dependent variable	0·535	0·535	0·423	0·423	0·294	0·294	0·556	0·556
Per decade trend as % of period mean	−6·4	−6·0	−5·9	−6·4	−6·5	−7·8	−9·4	−5·6
Pseudo-R^2	0·0011	0·018	0·0006	0·019	0·0005	0·017	0·004	0·026

Notes: (i) Results are for probit estimation, with coefficients reported at mean values and S.E. in parentheses. The coefficient for the foreign dummy is reported over the 0–1 range. (ii) The dependent variable is 0 if an inventor does not switch fields between two consecutive innovations. The dependent variable is 1 if the inventor does switch fields. Column headings define the field classification used to determine the dependent variable: "414" indicates the 414-category technological class definition of the U.S. Patent and Trademark Office; "36" and "6" refer to the aggregated measures defined by Hall *et al.* (2001). (iii) Specifications (1) through (6) consider "fast" innovators—only those consecutive patents with no more than 3 years between applications and with no more than 3 years delay between application and grant (see discussion in text). Specifications (7) and (8) consider all consecutive patents. (iv) Technological field controls are dummies for the 36 categories defined by Hall *et al.* (2001). The reported regressions use the technological field of the initial patent. Using the field of the second patent has no effect on the results. Running the regressions separately by technology category shows that the trends persist in six of the six categories using the broad technology classification of Hall *et al.* and 34 of the 36 categories using the narrow classification of Hall *et al.*, with significant trends in 20. (v) Assignee code controls are seven dummy variables that define who holds the rights to the patent. Table B.2 describes the assignee codes in further detail. The declining probability of field jump persists when the trend is examined within each assignment code, although the significance of the trend disappears in the rarer classifications.

TABLE 5

Mean differences across technological categories

6	Technological classification (Hall *et al.*, 2001) 36	Code	Age at first innovation Obs.	Age at first innovation Mean	Inventors per patent Obs.	Inventors per patent Mean	Probability of field jump Obs.	Probability of field jump Mean
Chemical (1)	Agriculture, Food, Textiles	11	12	31·1	18, 351	2·45	2461	0·48
	Coating	12	52	29·2	32, 820	2·24	4336	0·64
	Gas	13	17	29·9	10, 047	1·96	1692	0·59
	Organic Compounds	14	49	29·4	78, 188	2·57	8010	0·32
	Resins	15	43	29·4	74, 993	2·52	7695	0·37
	Miscellaneous—Chemical	19	329	29·3	214, 854	2·23	29, 721	0·43
	Entire category		502	29·4	429, 253	2·35	53, 915	0·43
Computers & Communications (2)	Communications	21	270	29·2	98, 046	1·99	15, 107	0·41
	Computer Hardware & Software	22	169	29·8	83, 094	2·26	10, 259	0·44
	Computer Peripherals	23	38	29·2	22, 809	2·37	2772	0·51
	Information Storage	24	45	29·0	43, 182	2·21	6778	0·40
	Entire category		522	29·4	247, 131	2·15	34, 916	0·42
Drugs & Medical (3)	Drugs	31	74	29·9	77, 210	2·91	7181	0·25
	Surgery & Medical Instruments	32	276	29·8	62, 192	1·86	12, 385	0·29
	Biotechnology	33	48	30·5	29, 638	2·75	2223	0·36
	Misc—Drugs & Medical	39	71	29·2	14, 356	1·66	3488	0·35
	Entire category		469	29·8	183, 396	2·43	25, 277	0·29
Electrical & Electronic (4)	Electrical Devices	41	116	29·2	65, 500	1·77	12, 817	0·48
	Electrical Lighting	42	88	29·6	33, 769	1·97	5739	0·43
	Measuring & Testing	43	117	29·1	62, 021	1·94	10, 083	0·51
	Nuclear & X-rays	44	56	29·5	32, 402	2·08	4681	0·50
	Power Systems	45	124	29·4	73, 849	1·94	13, 086	0·51
	Semiconductor Devices	46	51	29·3	47, 123	2·25	7207	0·34
	Misc—Electrical	49	103	29·1	52, 206	1·97	9004	0·51
	Entire category		655	29·3	366, 870	1·97	62, 617	0·48
Mechanical (5)	Materials Processing & Handling	51	243	29·4	108, 873	1·79	21, 821	0·48
	Metal Working	52	89	28·8	63, 669	2·12	10, 454	0·54
	Motors, Engines & Parts	53	86	29·4	78, 585	1·85	16, 221	0·41
	Optics	54	57	29·0	51, 102	2·15	8159	0·37
	Transportation	55	279	28·9	61, 501	1·66	12, 004	0·45
	Misc—Mechanical	59	458	29·1	103, 855	1·64	22, 513	0·49
	Entire category		1212	29·1	467, 585	1·84	91, 172	0·46
Others (6)	Agriculture, Husbandry, Food	61	248	29·1	44, 718	1·76	7644	0·40
	Amusement Devices	62	267	29·5	22, 227	1·41	4273	0·37
	Apparel & Textile	63	204	29·2	35, 001	1·57	7616	0·37
	Earth Working & Wells	64	98	29·7	29, 645	1·69	6599	0·36
	Furniture, House Fixtures	65	352	29·1	43, 499	1·42	9416	0·50
	Heating	66	55	30·0	28, 267	1·76	6065	0·48
	Pipes & Joints	67	46	29·2	18, 444	1·58	4448	0·61
	Receptacles	68	297	29·4	43, 353	1·51	10, 105	0·47
	Misc—Others	69	848	29·2	179, 925	1·74	35, 342	0·48
	Entire category		2415	29·3	445, 079	1·65	91, 508	0·46

Notes: (i) Age at first innovation includes observations of those innovators who appear after 1985 in the data set and between the ages of 23 and 33. Results are similar, with higher mean and even less variance, for 25- to 35-year-olds. (ii) Probability of field jump is probability of switching categories for solo innovators using 36-category measure. Obs., observation.

For a continuous measure of the quantity of knowledge, I use the logarithm of the number of nodes (*i.e.* patents) in the citation tree behind any patent.[9] As before, there is a truncation issue that needs to be considered: the data set does not contain citation information for patents issued before 1975, so we tend to see the recent part of the tree. The measure of underlying knowledge is then noisier the closer we are to 1975, and I therefore focus on cross-sections later in the time period. A second issue is that the average tree size and its variance grow extremely rapidly in the time window, which makes it difficult to compare data across cross-sections without a normalized measure. Two obvious normalizations are (1) a dummy for whether the tree size is greater than the within-period median, and (2) the difference from the within-period mean tree size, normalized by the within-period standard deviation. Results are reported using the latter definition, as it is informationally richer, though either method shows similar results.

Table 6 examines the relationship between team size and tree size in pooled cross-sections, with and without various controls. I add a quadratic term for the variation in team size to help capture evident curvature, and we see that team size rises at an increasing rate as the measure of knowledge depth increases. For innovations with larger citation trees, the rise in team size is particularly strong. With very deep knowledge trees, an increase of 1 S.D. in the tree size is associated with an average increase in team size of one person. The table shows that the cross-sectional relationship holds for domestic-source and foreign-source patents and when controlling for technological category, so that the variation appears both within fields and across them. Technological controls are perhaps best left out, however, since the variations in mean tree size across technological category may be equally of interest. Finally, we might be concerned that bigger teams simply have a greater propensity to cite, which results in larger trees. This concern proves unwarranted. Controlling for the variation in the direct citations made by each patent, we find that relationship actually strengthens. In fact, we see that bigger teams tend to cite *less*. This result gives us greater faith in the causative arrow implied by the regressions.

Next, we turn to the age at first innovation. Table 7 examines, in pooled cross-sections, the relationship between age and knowledge for those individuals for whom we can be confident that they are innovating for the first time (see discussion above). The general conclusion from the table is that we must work hard to find a relationship, and at its largest, it is very small. It is not robust to the specific age window, is reduced when controlling for the technological category, and disappears when controlling for the number of direct citations made. Taking a coefficient of 0·1 as the maximum estimate from the table, we find that an increase of 1 S.D. in the knowledge measure leads to a 0·1-year increase in age. This coefficient may be attenuated given that our proxy measure of knowledge is noisy, but I conclude that there is at most only a weak relationship between the amount of knowledge underlying a patent and the average age at first innovation.

Finally, Table 8 considers the relationship between the probability of field jump and the knowledge measure. The table shows a robust negative relationship: solo innovators are less likely to jump fields when their initial patent has a larger node count. If we identify a larger node count with a deeper area of knowledge, then this negative correlation is consistent with the idea that deeper areas of knowledge see more specialization. The results are robust to the inclusion of many controls, including controls for technological field, foreign or domestic source of the patent, and the time lag between the two patents. The results are also strengthened when examining

9. The distribution of the raw node count within cross-section is highly skewed—the mean is far above the median, so that upper tail outliers can dominate the analysis. I therefore use the natural log of the node count, which serves to contain the upper tail. A (loose) theoretical justification is knowledge depreciation: distant layers of the tree are less relevant to a patent than nearer layers, so there is a natural diminishing impact as nodes grow more distant. The diminishing impact of the large, distant layers, which dominate the node counts, is captured loosely by taking logs. Noting that the basic results are similar when we use the median-based measure of knowledge depth (a dummy for whether the raw node count is above or below the median, which is independent of any monotonic transform of the node count) we can be reasonably comfortable with the log measure.

TABLE 6

Inventors per patent versus tree size

	Dependent variable: inventors per patent						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Normalized variation in tree size	0·0849	0·0961	0·0995	0·120	0·133	0·107	0·152
	(0·0010)	(0·0010)	(0·0011)	(0·001)	(0·001)	(0·001)	(0·001)
Normalized variation in tree size, squared	0·0609	0·0545	0·0545	0·0341	0·0257	0·0356	0·0404
	(0·0007)	(0·0007)	(0·0007)	(0·0007)	(0·0009)	(0·0011)	(0·0009)
Foreign patent	—	0·446	0·442	0·420	U.S. only	Foreign only	0·371
		(0·002)	(0·002)	(0·002)			(0·003)
Normalized variation in direct citations made	—	—	−0·0094	—	—	—	—
			(0·0011)				
Technological field controls	—	—	—	Yes	Yes	Yes	Yes
Application year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	1,969,908	1,969,908	1,969,908	1,969,908	1,103,402	866,506	1,330,210
Period	1975–1999	1975–1999	1975–1999	1975–1999	1975–1999	1975–1999	1985–1999
Mean of dependent variable	2·02	2·02	2·02	2·02	1·82	2·27	2·13
R^2	0·026	0·050	0·050	0·100	0·090	0·083	0·079

Notes: (i) Regressions are ordinary least squares with S.E. in parentheses. Specifications (1) through (4) consider the entire universe of patents applied for between 1975 and 1999. Specifications (5) and (6) consider separately patents from domestic versus foreign sources. Specification (7) considers cross-sections from the later part of the time period. (ii) Normalized variation in tree size is the deviation from the year mean tree size, divided by the year S.D. in tree size. "Tree size" is the log of the number of nodes in the citations tree behind any patent. (iii) Normalized variation in direct citations made captures variation in the number of citations to prior art listed on a patent application. It is the deviation from the year mean number of citations, divided by the year S.E. in the number of citations. (iv) Technological field controls include dummies for each of 36-category measure of Hall *et al.* (v) The number of observations here is slightly smaller than for the time trend analysis in Table 2 because a few patents do not cite other U.S. patents; hence, no citation tree can be built; these patents are dropped from the analysis.

TABLE 7

Age versus tree size

	Dependent variable: age at application for first patent							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Normalized variation in tree size	−0·007	−0·005	0·114	0·084	0·059	0·097	0·113	0·030
	(0·032)	(0·036)	(0·035)	(0·040)	(0·043)	(0·030)	(0·046)	(0·026)
Team size	—	−0·054	—	−0·036	−0·038	−0·024	0·008	−0·029
		(0·027)		(0·030)	(0·030)	(0·025)	(0·035)	(0·019)
Normalized variation in direct citations made	—	—	—	—	0·064	—	—	—
					(0·044)			
Technological field controls	—	Yes	—	Yes	Yes	Yes	Yes	Yes
Application year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	6486	6486	5058	5058	5058	8434	3630	3588
Period	1985–1999	1985–1999	1985–1999	1985–1999	1985–1999	1975–1999	1985–1999	1985–1999
Age range	25–35	25–35	23–33	23–33	23–33	23–33	21–31	28–33
Mean of dependent variable	31·0	31·0	29·34	29·3	29·2	29·2	27·7	30·7
R^2	0·009	0·022	0·009	0·021	0·012	0·020	0·025	0·020

Notes: (i) Regressions are ordinary least squares, with S.E. in parentheses. All regressions observe only at those innovators for whom we have age data. Specifications (1) and (2) consider first innovations in the 25–35 age window. Specifications (3) through (6) consider innovators in the 23–33 age window. Specification (7) considers slightly younger innovators, and Specification (8) considers the latter half of the 23–33 age window. Specification (6) considers cross-sections pooled over the entire time period; the other specifications focus on the post-1985 period, for which we can be confident that we are witnessing an innovator's first patent. (ii) Normalized variation in tree size is the deviation from the year mean tree size, divided by the year S.E. in tree size. Tree size is the log of the number of nodes in the citations tree behind any patent. (iii) Normalized variation in direct citations made captures variation in the number of citations to prior art listed on a patent application. It is the deviation from the year mean number of citations, divided by the year S.E. in the number of citations. (iv) The number of observations here is slightly smaller than for the time trend analysis in Table 3 because a few patents do not cite other U.S. patents; hence, no citation tree can be built; these patents are dropped from the analysis. (v) Technological field controls include dummies for each of 36-category measure of Hall *et al.*

TABLE 8

Field jump versus tree size

	Dependent variable: probability of switching technological field					
	(1)	(2)	(3)	(4)	(5)	(6)
Normalized variation in tree size	−0·0072	−0·0074	−0·0059	−0·0095	−0·0144	−0·0184
	(0·0008)	(0·0008)	(0·0008)	(0·0009)	(0·0012)	(0·0017)
Foreign patent	—	−0·0125	−0·0108	−0·0129	−0·0135	0·0032
		(0·0018)	(0·0018)	(0·0018)	(0·0023)	(0·0032)
Time between applications	—	—	0·0226	0·0232	0·0215	0·0143
			(0·0004)	(0·0004)	(0·0012)	(0·0017)
Technological field controls (first patent)	—	—	—	Yes	Yes	Yes
Application year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	353,762	353,762	353,762	353,762	212,274	110,511
Period	1975–1999	1975–1999	1975–1999	1975–1999	1975–1993	1985–1993
Mean of dependent variable	0·551	0·551	0·551	0·551	0·536	0·520
Pseudo-R^2	0·0039	0·0039	0·0117	0·0251	0·0171	0·0159

Notes: (i) Results are for probit estimation, with coefficients reported at mean values and S.E. in parentheses. The coefficient for the foreign dummy is reported over the 0–1 range. Only solo inventors are considered. Specifications (1) through (4) consider the entire set of solo inventors. Specification (5) considers only those solo inventors who meet the criteria in Specifications (1) through (6) in Table 4 (to help control for any truncation bias in the specialization measure—see the discussion of Table 4 in the text). Specification (6) considers the same data as Specification (5) but only examines cross-sections in the later part of the time period. (ii) The dependent variable is 0 if an inventor does not switch fields between two consecutive innovations. The field is defined using the 414-category technological class definition of the U.S. Patent and Trademark Office. (iii) Normalized variation in tree size is the deviation from the year mean tree size, divided by the year S.E. in tree size. Tree size is the log of the number of nodes in the citations tree behind any patent. (iv) Technological field controls include dummies for each of 36-category measure of Hall *et al.*

cross-sections later in the time period, where the citations trees capture more historical information and may be less noisy measures of the underlying knowledge.

To summarize, we have presented six facts about innovators. Using the measures defined previously, we find that specialization and teamwork appear to increase with time and are also greater, in cross-section, in deeper areas of knowledge. Meanwhile, the average age at first innovation is increasing with time, like specialization and teamwork, but shows little variation with the depth of knowledge in cross-section. The following section presents a model, building on the burden of knowledge idea, which (a) shows how these behaviours can all emerge in equilibrium and (b) clarifies the growth implications. Further, related evidence from existing literature is discussed in Section 4.

3. THE MODEL

The model considers innovator behaviour along a balanced growth path. Building on foundations of existing growth models, I analyse a structure with two sectors: a production sector where competitive firms produce a homogenous output good and an innovation sector where innovators produce productivity-enhancing ideas. The novelty of the model lies in the innovators' choice problem. Innovators undertake costly human capital investments to bring themselves to the knowledge frontier. Innovators weigh the costs and benefits of gaining particular forms of expertise, decisions that will be balanced differently by different cohorts as the economy evolves and balanced differently in different areas of application. The model ties together the facts of Section 2 on the basis of these educational decisions and shows how the burden of knowledge interacts with other forces in determining the steady-state growth rate.

Section 3.1 describes the production sector and Section 3.2 defines individuals' life cycles and preferences. Sections 3.3 and 3.4 focus on innovators. The first describes the knowledge space and the cost of education. The second considers the innovation process and the value of ideas. Section 3.5 defines individuals' equilibrium choices, and Section 3.6 analyses educational decisions and growth along a balanced growth path. Proofs are presented in Appendix A.

3.1. *The production sector*

There is a continuum of productive ideas of measure $N(t) > 0$ at each time $t \in \mathbb{R}$. Let each idea $k \in [0, N(t)]$ make a productivity contribution denoted $\gamma(k) > 0$. Define $X(t) = \int_0^{N(t)} \gamma(k)dk$ as the collective productivity contribution of all existing ideas at time t.

Let there be a homogenous output good produced by competitive firms at each time t. The price of the good is normalized to 1 at each point in time. A firm hires an amount of labour, $l(t)$, producing output $y(t) = X(t)l(t)$ if all existing ideas $k \in [0, N(t)]$ are employed by the firm.[10]

Firms pay workers a competitive wage, $w(t)$, and also make royalty payments. The royalty payment per production worker is $r(k,t)$ for idea k at time t, and the total royalty payments per production worker are $r(t) = \int_0^{N(t)} r(k,t)dk$ when all existing ideas $k \in [0, N(t)]$ are employed by the firm. Profits are $\pi(t) = (X(t) - r(t) - w(t))l(t)$ when all existing ideas are employed. Ideas receive patent protection for a finite number of years $z > 0$. It is straightforward to show that the monopolist owner of an unexpired patent can charge a royalty per worker $r(k,t) = \gamma(k)$, and competitive firms will be just willing to pay this fee.[11] Meanwhile, $r(k,t) = 0$ for expired patents, which are freely available. Hence, firms employ all available ideas, paying royalties on all unexpired patents totalling $r(t) = X(t) - X(t-z)$ per production worker. Total output in the economy is:

$$Y(t) = X(t)L_Y(t), \tag{1}$$

where $L_Y(t)$ is the total mass of production workers. Competitive firms earn zero profits, so that $w(t) = X(t) - r(t)$, and the wage paid to a production worker is therefore:

$$w(t) = X(t-z). \tag{2}$$

3.2. *Workers and preferences*

There is a continuum of workers of measure $L(t) > 0$ in the economy at time t. This population grows at rate $g_L > 0$. Individuals have a common hazard rate ϕ of death. Individuals are risk-neutral, with expected utility for an individual i defined by:

$$U^i(\tau) = \int_\tau^\infty c^i(\tau,t)e^{-\phi(t-\tau)}dt, \tag{3}$$

where $c^i(\tau,t)$ is the consumption at time t of an individual i born at time τ.

10. Firms use the whole set of existing ideas rather than just the latest idea. That is, we use a "horizontal" model of innovation, where ideas accumulate rather than become obsolete (see, *e.g.* Barro and Sala-i-Martin, 1995, for a review).

11. This production set-up follows closely on Arrow (1962) and Nordhaus (1969). The maximization problem of a firm can be written explicitly as follows. Define $\widetilde{X}(t) = \int_0^{N(t)} \gamma(k)I(k,t)dk$ and $\widetilde{r}(t) = \int_0^{N(t)} r(k,t)I(k,t)dk$ where $I(k,t) = 1$ if the firm employs idea k at time t and $I(k,t) = 0$ otherwise. The firm's profit is $\widetilde{\pi}(t) = (\widetilde{X}(t) - \widetilde{r}(t) - w(t))l(t) = (\int_0^{N(t)}[\gamma(k) - r(k,t)]I(k,t)dk - w(t))l(t)$. To maximize profits, the firm chooses which ideas k to employ, setting $I(k,t) = 1$ when $\gamma(k) \geq r(k,t)$ and $I(k,t) = 0$ otherwise. The holder of a patent sets $r(k,t) = \gamma(k)$ when the patent is valid, while $r(k,t) = 0$ when the patent has expired. The firm thus sets $I(k,t) = 1$ for all $k \in [0, N(t)]$. Firms produce with productivity $\widetilde{X}(t) = X(t)$, paying royalties on all unexpired patents totalling $\widetilde{r}(t) = X(t) - X(t-z)$ per worker.

I assume that individuals are born without assets and supply a unit of labour inelastically at all points over their lifetime. Following standard models of finite horizons (*e.g.* Blanchard, 1985), we allow for competitive life insurance and annuity firms so that loans are secured by life insurance and assets held as annuities; thus, workers do not die in debt or with positive assets. In the absence of physical capital, the discount rate in this model is simply ϕ, the hazard rate of death.[12] From the standard intertemporal budget constraint, the individual's utility is equivalent to the present value of his or her expected lifetime non-interest income.

The individual maximizes lifetime income through the choice of career. This is a permanent decision made at birth. In particular, the individual may become (i) a wage worker or (ii) an innovator. Wage workers require no education, and the expected present value of their lifetime income is the discounted flow of the wage payments, $w(t)$, they receive. For a wage worker born at time τ,

$$U^{\text{wage}}(\tau) = \int_{\tau}^{\infty} w(t)e^{-\phi(t-\tau)}dt. \tag{4}$$

If an individual chooses to be an innovator, then he or she must further choose a specific field of expertise, represented by a vector $\boldsymbol{\omega^i}$. The innovator pays an immediate educational cost at birth, $E(\boldsymbol{\omega^i}, \tau)$, to bring himself or herself to the frontier of knowledge in the chosen field. He or she earns an expected flow of income, $v(\boldsymbol{\omega^i}, t)$, throughout their life that comes from royalties on any innovations he or she produces. The expected present value of lifetime income for an innovator born at time τ is:

$$U^{\text{R\&D}}(\boldsymbol{\omega^i}, \tau) = \int_{\tau}^{\infty} v(\boldsymbol{\omega^i}, t)e^{-\phi(t-\tau)}dt - E(\boldsymbol{\omega^i}, \tau). \tag{5}$$

The structure of the innovator's educational decision, $\boldsymbol{\omega^i}$, and the functional forms of $E(\boldsymbol{\omega^i}, \tau)$ and $v(\boldsymbol{\omega^i}, t)$ are defined in the following subsections.

3.3. *Knowledge and education*

Let knowledge be organized as follows. First, there are "areas of application". Second, there is "foundational knowledge" underneath an area of application. For example, one application area could be airplanes, building on foundational knowledge of fluid mechanics, thermodynamics, and material science. Another application area could be drugs, building on knowledge of immunology, protein synthesis, and bioinformatics. The amount of foundational knowledge may differ for different areas of application.[13]

Formally, let there be J areas of application, indexed $j \in \{1, 2, \ldots, J\}$, where J is finite. Within each area of application, there is a set of types of foundational knowledge arranged around a circle of unit circumference. We denote each such circle Λ_j and a point on such a circle as s_j. The measure of foundational knowledge in an application area is denoted $D_j(t)$, as shown in Figure 3. As helpful nomenclature, we refer to Λ_j as a "circle of knowledge" and the measure $D_j(t)$ the "depth of knowledge".

12. The riskless rate of return is zero in the absence of physical capital; the discount rate exists purely to cover the possibility of death. In particular, the insurance premium to secure loans and the rate of return on annuities are both equivalent to the hazard rate of death under the zero-profit condition for insurance and annuity firms. For example, an annuity firm pays a stream v while you live in exchange for a dollar invested today. Expected profits for the annuity firm are $1 - v/\phi$. The zero-profit condition then requires $v = \phi$.

13. For simplicity, we assume that all areas of application are used in the production of the homogenous output good. One could alternatively allow for multiple types of output goods based on different areas of application, but such an extension would distract from the core mechanism of the model and is thus left aside.

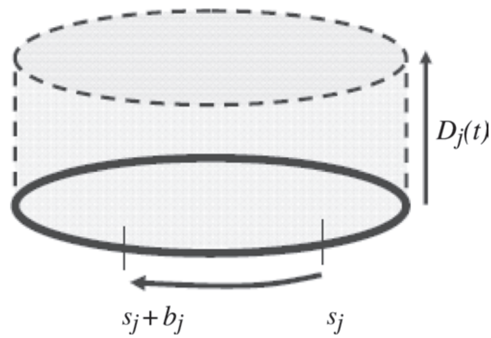

FIGURE 3
A circle of knowledge

A prospective innovator chooses an area of expertise, a vector $\boldsymbol{\omega} = (j, s_j, b_j)$, which defines (1) an application area $j \in \{1, 2, \ldots, J\}$; (2) a point, $s_j \in \Lambda_j$, in the set of foundational knowledge types underlying application area j; and (3) a certain distance, $b_j \in [0, \infty)$, measured clockwise of s_j.[14] To ease notation, we have dropped the superscript i in the vector $\boldsymbol{\omega}$, and ask the reader to recall that ω is a choice made by an individual i.

For an innovator born at time τ, the amount of knowledge the innovator acquires is the individual's chosen breadth of expertise, b_j, multiplied by the prevailing depth of knowledge, $D_j(\tau)$. The educational cost of acquiring this information is:

$$E(\boldsymbol{\omega}, \tau) = (b_j D_j(\tau))^{\varepsilon}, \tag{6}$$

where $\varepsilon > 0$, which says that learning more requires a greater amount of education.

The depth of knowledge may differ across different application areas and may evolve with time. In particular,

$$D_j(t) = \hat{D}_j X(t)^{\delta} \tag{7}$$

where $\hat{D}_j > 0$ is specific to the application area. Thus, the amount of foundational knowledge for airplanes may be relatively large (high $\hat{D}_j$) compared to amusement devices (low $\hat{D}_j$). With $\delta > 0$, the depth of knowledge increases over time as the productivity level of the economy advances.

3.4. *Innovation*

Once educated, innovators begin to receive innovative ideas. The total stock of ideas at time t is $N(t)$. Let each idea (*i.e.* each unit mass of ideas) add to productivity by an amount γ, so that productivity evolves as $X(t) = \gamma N(t)$ when the total stock of ideas is employed.[15] Recalling that

14. We allow b_j to take values greater than 1—that is for an innovator's expertise to wrap around the circle multiple times. One can imagine that innovators gain further educational value by covering the same foundational knowledge again; for example, re-reading material creates better understanding than one's first read. This assumption is largely made for technical reasons, however, to avoid dealing with corner solutions where choices of b_j are capped at some finite maximum value. Corner solutions can be handled in a variation of the model but are awkward and add no important insights.

15. One could alternatively allow the size of ideas to grow or decline with time, or allow the size of ideas to be functions of educational choices. Such specifications would have no substantive effect on the analysis. See Jones (2005*b*) and the discussion in Section 4.

an idea can be licensed to $L_Y(t)$ workers and patents last for z years, the lump sum value of an idea is:

$$V(t) = \gamma \int_t^{t+z} L_Y(\widetilde{t})d\widetilde{t}. \tag{8}$$

Note that patents do not expire on the death of an innovator.[16] Like any asset, the innovator prefers to hold a patent as an annuity, selling a patent to a competitive annuity firm in exchange for an annuity $\phi V(t)$. For the innovator, the present value of this annuity at time t is $V(t)$.[17]

3.4.1. Inspiration. Ideas comes to an innovator at rate $\lambda(\boldsymbol{\omega}, t)$. This arrival rate depends in part on an innovator's educational decision, $\boldsymbol{\omega}$, which is fixed over an innovator's lifetime, and in part on the overall state of the economy, which evolves with time. In particular,

$$\lambda(\boldsymbol{\omega}, t) = \hat{A}_j X(t)^{\chi} L_j(t, s_j)^{-\sigma} b_j^{\beta}, \tag{9}$$

where $L_j(t, s_j)$ is the mass of living innovators at time t who have chosen location s_j, while $\hat{A}_j$ represents application-area-specific research opportunities.

This reduced-form specification captures several key ideas. The parameter χ represents the impact of the current state of technology on an innovator's creative output. It incorporates the standard ideas in the literature alluded to in the introduction: fishing-out hypotheses whereby innovators' productivity falls as the state of knowledge advances ($\chi < 0$), and rising technological opportunity whereby an improving state of knowledge makes innovators more productive ($\chi > 0$). The term $\hat{A}_j > 0$ meanwhile allows for technological opportunities to vary across application areas—for some areas to be relatively "hot" or "cold".

The parameter σ represents the impact of crowding on the frequency of an innovator's ideas. I assume $0 < \sigma < 1$, following standard arguments where innovators partly duplicate each other's work. A greater density of workers in the same specialty increases duplication, reducing the rate at which a specific individual produces a novel idea.[18]

The final parameter, β, represents the impact of the breadth of expertise. We assume $\beta > 0$, which says simply that broader foundational knowledge increases one's productivity. This is natural if, for example, access to a broader set of available knowledge—facts, theories, methods—creates better combinatorial possibilities for one's creativity, along the lines of Weitzman (1998), making the innovator more productive.[19]

3.4.2. Implementation. Ideas are implemented by pooling requisite foundational knowledge. Implementation thus involves the formation of teams. This process operates under simplifying assumptions as follows.

16. This is a realistic feature of the model: in the real world, patent rights are assignable and patents do not expire on the death of an inventor.

17. If the innovator did not have access to a competitive financial market that pays the innovator the lump sum value of the patent (or an equivalent annuity) in exchange for the patent rights, then the value of the patent to the innovator would need to reflect the possibility that the innovator dies before the patent rights expire, in which case $V(t) = \gamma \int_t^{t+z} L_Y(\widetilde{t})e^{-\phi(\widetilde{t}-t)}d\widetilde{t}$. This variation will have no impact on the main results of the model.

18. An alternative formulation of equation (9), where individuals crowd over an interval of knowledge rather than a point of knowledge, can explain the six stylized facts of Section 2 along the same lines as this model but is less tractable.

19. There are many other mechanisms through which broader expertise would enhance an innovator's productivity. For example, a more broadly expert innovator may better evaluate the expected impact and feasibility of his or her ideas. He or she will better select towards high value, successful lines of inquiry, and therefore achieve greater returns. Furthermore, if assembling teams is costly, innovators will be unwilling to form large teams. More broadly, expert innovators can rely less on large teams for the implementation of their ideas, making their ideas less costly to implement.

First, it is assumed that implementation requires all types of foundational knowledge in the given area of application. Formally, define ideas in application area j as *implementable* if for each $s_j \in \Lambda_j$, there exists an individual l in application area j such that s_j lies in the arc between s_j^l and $s_j^l + b_j^l$. Let $I_j(t)$ be an indicator function equal to 1 if the idea is implementable and 0 otherwise.

Second, we assume that the innovator with the idea claims its rents. One may imagine that innovators work in firms, pooling knowledge in application area j, where the wage paid to each innovative worker as the value of his or her ideas. Alternatively, one may think of the innovator with an idea as a monopolist vis-à-vis potential teammates so that the inspired innovator extracts all profits from the project. We abstract from costs in team formation or operation, so that all ideas are profitable with lump sum value $V(t)$.

The expected flow of income to an innovator, $v(\boldsymbol{\omega}, t)$, is then the probability an idea arrives and is implementable times the lump sum value of the idea.[20] Hence,

$$v(\boldsymbol{\omega}, t) = \lambda(\boldsymbol{\omega}, t) I_j(t) V(t). \tag{10}$$

Finally, we consider below the size of teams. To simplify this analysis, we assume that an inspired innovator forms teams within his or her own cohort if possible and assembles the minimum number of people necessary to implement the idea.

3.5. *Equilibrium career choices*

In equilibrium, a player cannot make a different choice of career and be better off. For an individual born at any time τ, the decision to become a wage worker requires that

$$U^{\text{wage}}(\tau) \geq U^{\text{R\&D}}(\boldsymbol{\omega}', \tau) \quad \forall \boldsymbol{\omega}',$$

so that wage workers would not strictly prefer to be R&D workers. Similarly, for an individual born at any time τ, the decision to become an R&D worker with educational choice $\boldsymbol{\omega}$ requires that

$$U^{\text{R\&D}}(\boldsymbol{\omega}, \tau) \geq U^{\text{R\&D}}(\boldsymbol{\omega}', \tau) \quad \forall \boldsymbol{\omega}',$$
$$U^{\text{R\&D}}(\boldsymbol{\omega}, \tau) \geq U^{wage}(\tau),$$

so that R&D workers of type $\boldsymbol{\omega}$ would not strictly prefer to be R&D workers of a different type or wage workers.

3.5.1. Balanced growth path. We focus on equilibrium career decisions along a balanced growth path. A balanced growth path is defined such that the growth rate in productivity, $g = \dot{X}(t)/X(t)$, is constant over time and the labour allocations $L_Y(t)$ and $L_j(t,s)\ \forall j, s$ grow with

20. One can also consider rent sharing among teammates, which adds considerable complexity. With rent-sharing, equilibrium income flow will still take the form of equation (10). This follows because innovators in the same cohort earn the same income in equilibrium, which must then be the per capita rate of idea arrival times the value of ideas. At the same time, rent sharing can create an inefficiency should innovators expand their expertise not only to improve their creative output but also to claim greater royalty shares from their teammates. While rent sharing can thus affect the benefits of breadth, the basic idea that the burden of knowledge raises the cost of breadth, provoking increased specialization and teamwork, will be robust under a wide variety of rent-sharing arrangements. One might also consider many other possible frictions and inefficiencies in team formation. The model featured imagines that such frictions and incentive issues are solved, allowing us to focus on a benchmark outcome. Also, see Jones (2008) for a model that features the intersection between educational decisions and frictions in team formation.

time at the population growth rate g_L. The existence of a balanced growth path is established in the analysis below.

We analyse the balanced growth path under three parametric restrictions that are assumed throughout the following analysis.

Assumption 1. $\beta < \varepsilon$.

This assumption is necessary for an innovator's optimal breadth of expertise, the choice b_j, to be an interior maximum.

Assumption 2. $\chi - \beta(\delta - \frac{1}{\varepsilon}) < 1$.

This assumption is necessary for the existence of a constant productivity growth rate g.

Assumption 3. $\phi > \max\{g, [1+\beta(\delta - \frac{1}{\varepsilon})]g\}$, *where* $g = \frac{1-\sigma}{1-\chi+\beta(\delta-\frac{1}{\varepsilon})} g_L$.

This assumption is necessary for an individual's lifetime income to be finite.[21]

3.6. *Analysis*

Production workers receive a competitive wage $w(t) = X(t-z)$ as shown in equation (2). Along a balanced growth path, $X(t-z)$ grows at rate g so that, from equation (4), a production worker earns lifetime income:

$$U^{\text{wage}}(\tau) = \frac{X(\tau - z)}{\phi - g}. \tag{11}$$

where we require $\phi > g$ for finite lifetime income.[22]

The innovator, meanwhile, makes an educational choice to maximize lifetime income. With the objective function (5) and the definitions in equations (6), (9), and (10), the innovator's problem is:

$$\max_{\boldsymbol{\omega}=(j,s_j,b_j)} \int_{\tau}^{\infty} \hat{A}_j X(t)^{\chi} L_j(t,s_j)^{-\sigma} b_j^{\beta} I_j(t) V(t) e^{-\phi(t-\tau)} dt - (b_j D_j(\tau))^{\varepsilon}. \tag{12}$$

The first result regarding career choice establishes the useful property of income equivalence between innovators and wage workers.

Lemma 1. *Along a balanced growth path* $U^{\text{R\&D}}(\boldsymbol{\omega}, \tau) = U^{\text{wage}}(\tau)$ *for any equilibrium choice* $\boldsymbol{\omega}$ *and any cohort* τ.

This income equivalence result rules out corner solutions where all individuals choose to be wage workers or all choose to be innovators. It follows naturally in the set-up of the model. Wage workers are needed to create a market for innovations, and innovators are too productive when rare to fail to exist. Along a balanced growth path, masses of wage workers and innovators are all growing, so that individuals actively choose both broad types of careers in every cohort, and hence, their income must be equivalent in equilibrium.

The next results further define innovator behaviour, building on the choice of $\boldsymbol{\omega} = (j, s_j, b_j)$.

21. The death rate ϕ is the discount rate. One could add a pure rate of time preference to the model, in addition to the death rate ϕ, which would raise the discount rate and allow lifetime income to be bounded under higher growth rates.

22. It is demonstrated in the proof of Proposition 2 that finite income for wage workers follows from Assumption 3 along a balanced growth path.

Proposition 1. *Along a balanced growth path:*

(i) $I_j(t) = 1$ *for all* j, t.
(ii) $L_j(t, s') = L_j(t, s'')$ *for all* s', s'' *in an area of application* j.
(iii) $E(\boldsymbol{\omega}, \tau)/U^{\mathrm{R\&D}}(\boldsymbol{\omega}, \tau) = \frac{\beta}{\varepsilon - \beta}$, *where* $\beta < \varepsilon$.

Result (i) says that innovators exist with sufficient expertise to implement any idea in any application area. This follows because duplication is costly so that innovators seek to avoid crowding. In particular, with $\sigma > 0$, any area of application with no active innovators becomes too tempting to ignore—an innovator would always deviate to such an area. Result (ii) follows from similar reasoning. It says that innovators spread evenly within a given application area. This follows because, within an application area, there are no costs or benefits of a particular location s, except the relative density of innovators there. Hence, with $\sigma > 0$, innovators avoid crowding and array themselves evenly. For clarity, we denote the labour allocation $L_j(t, s)$ as $L_j^*(t)$ in equilibrium to emphasize that it is independent of s in a given area of application j. The total mass of innovators at time t is then $L_R^*(t) = \sum_{j=1}^{J} L_j^*(t)$, and the total mass of wage workers is $L_Y^*(t) = L(t) - L_R^*(t)$.

Result (iii) is less obvious and more powerful. It says that the ratio between educational expenditure and lifetime income is constant, regardless of the innovator cohort or particular equilibrium area of expertise. This follows from the choice of b_j, which equates the marginal cost and benefit of breadth. Generally, if we view $D_j(\tau)$ as the "price" of breadth, then an increased price results in decreased breadth, offsetting the rise in total educational cost. In this model, price and quantity are traded off exactly so that educational cost is a constant fraction of lifetime income. This type of result should be familiar from Cobb–Douglas specifications, which feature constant expenditure shares.[23] This result requires that a choice b_j represents an interior maximum, so that the marginal benefits and costs of breadth are equated. This is guaranteed as long as $\beta < \varepsilon$ (Assumption 1), as shown in Appendix A.

Result (iii) is a key property of the equilibrium from which other results follow. As a first example, recall that $U^{\mathrm{R\&D}}(\boldsymbol{\omega}, \tau) = U^{\mathrm{wage}}(\tau)$ in equilibrium, so that innovators' income is independent of the particular equilibrium choice $\boldsymbol{\omega}$. It then follows directly from result (iii) that $E(\boldsymbol{\omega}, \tau)$ must likewise be independent of the particular equilibrium choice $\boldsymbol{\omega}$. This result is encapsulated as part of the following corollary.

Corollary 1. *Total education:*

(i) *(Cross-section)* $E(\boldsymbol{\omega}, \tau) = E(\boldsymbol{\omega}', \tau)$ *for any two equilibrium choices* $\boldsymbol{\omega}$, $\boldsymbol{\omega}'$ *made by individuals in the same cohort* τ.
(ii) *(Time series)* $E(\boldsymbol{\omega}, \tau)$ *grows across cohorts at rate* $g_E = g$.

These results inform the two key empirical facts regarding educational attainment from Section 2. Result (i) says that innovators in the same cohort choose the same amount of education across different areas of application. What is particularly surprising is that this result holds even though some areas may feature a greater difficulty in reaching the frontier of knowledge (higher $\hat{D}_j$) and some areas may be "hotter" than others, featuring more innovative opportunities (higher $\hat{A}_j$). This uniformity of education is possible through the endogenous allocation of innovators to different careers. Innovators allocate themselves across application areas to neutralize income differences (and hence educational differences) using differences in the degree of congestion to offset variation in technological opportunities or educational burden.

23. The technical basis for this type of result lies in isoelasticity. In particular, innovator output is isoelastic in breadth (just as output is isoleastic to the inputs in a Cobb–Douglas specification). Isoelasticity drives the constant expenditure share.

Result (ii) follows directly along the balanced growth path, where income is growing at rate g and hence education is too, maintaining a constant ratio as dictated by Proposition 1, part (iii). The model thus can match two key empirical facts of Section 2: common educational attainment across widely different areas of application, yet growing education over time.

Denote the common educational attainment $E(\boldsymbol{\omega}, \tau)$ within a cohort τ as $E^*(\tau)$. This equivalence of education in turn has direct implications for the breadth of expertise. Rearranging equation (6), we see that:

$$b_j = E(\boldsymbol{\omega}, \tau)^{1/\varepsilon}/D_j(\tau). \tag{13}$$

Common $E(\boldsymbol{\omega}, \tau) = E^*(\tau)$ within a cohort then implies that the equilibrium breadth of expertise will differ only by area of application j and cohort τ. In particular, denoting the equilibrium choice of breadth as $b_j^*(\tau)$ and the growth rate in $b_j^*(\tau)$ across successive cohorts as $g_{b_j^*}$, we find the following results.

Corollary 2. *Breadth of expertise:*

(i) *(Cross-section)* $b_j^*(\tau)/b_{j'}^*(\tau) = \hat{D}_{j'}/\hat{D}_j$.
(ii) *(Time series)* $g_{b_j^*} = (1/\varepsilon - \delta)\, g$*, so that* $g_{b_j^*} < 0$ *iff* $\delta > 1/\varepsilon$.

The first result says that innovators in areas with deeper knowledge choose narrower expertise. This follows naturally from common $E^*(\tau)$—where the depth of knowledge is higher and the breadth of expertise falls (see equation (13)). Interestingly, although field-specific technological opportunities influence the marginal benefit of breadth (see the $\hat{A}_j$ term in equation (12)), the endogenous labour allocation across fields neutralizes this effect, so that the relative breadth of expertise across fields is independent in equilibrium of how valuable knowledge is, and is determined solely from the cost side.

The second result tells how the breadth of expertise evolves along the growth path. From equation (13), the evolution of specialization across cohorts is a race between growing educational attainment, $E^*(\tau)$, and a growing distance to the knowledge frontier, $D_j(\tau)$. Only when the distance to the frontier is growing at a sufficient rate (high enough δ) will workers become more specialized even as they invest more in education.

The model thus can also match two further empirical facts of Section 2 regarding specialization: greater specialization in areas with deeper underlying knowledge and increasing specialization over time. Moreover, increasing specialization—despite increasing educational attainment—is directly associated in the model with increasing depth of knowledge along the growth path.

Finally, a simple, related outcome regards teamwork. In the model, innovation requires expertise over the whole set of knowledge underlying a given area of application.[24] Hence, teamwork is required when an individual innovator does not cover the entire circle of knowledge. With an equilibrium decision $b_j^*(\tau)$, the team size in a given cohort and application area is:[25]

$$\text{team}_j(\tau) = \begin{cases} 1 & b_j^*(\tau) > 1 \\ \lceil 1/b_j^*(\tau) \rceil & b_j^*(\tau) \le 1, \end{cases} \tag{14}$$

where $\lceil x \rceil$ is the ceiling function; that is, $\lceil x \rceil$ is the least integer $\ge x$. The following corollary thus follows directly from the last.

24. See Jones (2005*b*) for a model where implementation of ideas need not cover the entire set of knowledge in a given area of application. That model details a more general set of conditions under which greater teamwork follows from increased specialization.

25. Recall from Section 3.4.2 that teams are formed (i) within the same cohort when possible and (ii) with the minimum number of necessary teammates.

Corollary 3. *Teamwork*

(i) *(Cross-section)* $\text{team}_j(\tau) \geq \text{team}_{j'}(\tau)$ *iff* $\hat{D}_j \geq \hat{D}_{j'}$.
(ii) *(Time series)* $\text{team}_j(\tau) \geq \text{team}_j(\tau')$ *for any* $\tau > \tau'$ *iff* $\delta > 1/\varepsilon$.

The model therefore identifies greater teamwork in cross-section with deeper areas of knowledge and identifies increased teamwork over time with a rising burden of knowledge. Collectively, Corollaries 1, 2, and 3 show how innovator behaviour varies across fields and evolves as the economy grows, providing a unified and consistent interpretation for the six facts presented in Section 2.

A key mechanism in pinning down innovator behaviour is their choice of application area. This endogenous choice allows the equalization of lifetime income, which in turn allows the model to pin down educational attainment and other behaviours as shown previously. It is instructive to show explicitly the resulting allocation of labour across application areas.

Corollary 4. *Along a balanced growth path, the ratio of labour allocations in different application areas in the R&D sector is a constant where:*

$$\frac{L_j^*(t)}{L_{j'}^*(t)} = \left[\frac{\hat{A}_j}{\hat{A}_{j'}}\left(\frac{\hat{D}_{j'}}{\hat{D}_j}\right)^{\beta}\right]^{1/\sigma}. \tag{15}$$

We see that innovators are attracted to hot application areas (high $\hat{A}_j$) and areas with low learning costs (low $\hat{D}_j$). The interesting consequence is that while choice of application area is influenced by these cross-field variations, educational attainment does not vary across areas (Corollary 1). Meanwhile, breadth of expertise does vary with knowledge depth (Corollary 2). The endogenous labour allocation thus helps neutralize educational attainment but not specialization, allowing the model to unify the facts of Section 2.

3.6.1. Steady-state growth. We now consider the implications of the knowledge burden mechanism for aggregate growth. Growth comes from the summation of contributions from all innovators alive at a given moment. If there are $L_R^*(t)$ innovators active in equilibrium at time t and these innovators raise productivity in the economy on average at rate $\bar{\theta}(t)$, then productivity increases per unit of time are simply $\dot{X}(t) = \bar{\theta}(t)L_R^*(t)$. The growth rate of productivity is then:

$$g = \frac{\bar{\theta}(t)L_R^*(t)}{X(t)}. \tag{16}$$

Calculating innovators' average contributions, $\bar{\theta}(t)$, appears complicated because innovators are active in different areas of application with unique innovative opportunities and knowledge depth, and innovators come from different cohorts. However, aggregating innovators' contributions is simplified by the following result. In equilibrium, innovators in the same cohort add to productivity at the same rate, regardless of their area of application. The intuition builds from the results above: once individuals in the same cohort have equivalent $U^{\text{R\&D}}(\boldsymbol{\omega}, \tau)$ and equivalent $E(\boldsymbol{\omega}, \tau)$ in equilibrium, their expected gross income $(U^{\text{R\&D}}(\boldsymbol{\omega}, \tau) + E(\boldsymbol{\omega}, \tau))$ from innovation, and hence, their productivity contributions must also be equivalent. This property, which is shown formally in the proof of the following proposition, allows g to be determined as a simple function of exogenous parameters.

Proposition 2. *Along a balanced growth path,*

$$g = \frac{1-\sigma}{1-\chi+\beta(\delta-\frac{1}{\varepsilon})} g_L, \tag{17}$$

where $\chi-\beta(\delta-\frac{1}{\varepsilon})<1$. *There is a unique balanced growth path in equilibrium, with the constant growth rate* g *given in equation* (17) *and a set of labour allocations* $\{L_1^*(t),\ldots,L_J^*(t),L_Y^*(t)\}$ *where each labour allocation grows at rate* g_L.

The expression (17) defines the growth rate as the outcome of several important forces, marrying the knowledge burden mechanism with several ideas in the existing growth literature. First, the parameter χ, as discussed previously, represents standard ideas in the growth literature, whereby the productivity of innovators may increase as they gain access to new technologies and new ideas ($\chi>0$) or decrease if innovators are fishing out ideas ($\chi<0$). The larger χ—the greater the value of knowledge to further innovation—the greater the growth rate, as is seen in equation (17).

Second, the term $\beta(\delta-\frac{1}{\varepsilon})$ captures the implications of the burden of knowledge. The term $(\delta-\frac{1}{\varepsilon})$ is recognized from Corollary 2. With $\delta>1/\varepsilon$, innovators choose increasing specialization as the economy evolves, and we witness the "death of the Renaissance Man" along the growth path. The impact of narrowing expertise on growth will be large or small depending on the value of β, which defines the sensitivity of innovators' productivity to their breadth of expertise.[26]

Expression (17) also shows that the model eliminates scale effects. The productivity growth rate is constant despite an exponentially increasing scale of research effort, with the number of researchers growing at the population growth rate, g_L. In the model, growing population provides both the motive—increasing market size—and the means—more minds—for innovative effort to grow at an exponential rate in equilibrium, even if innovation is getting harder per person.

From a growth point of view, the burden of knowledge parameters $\beta(\delta-\frac{1}{\varepsilon})$ are seen to act similarly in equation (17) to the parameter χ that captures any fishing-out effect. Two interpretations of the burden of knowledge mechanism are then possible. First, the burden of knowledge mechanism can be seen as a micro-foundation for fishing-out-type effects on growth without literally believing that ideas are being fished out. Alternatively, if one is convinced that a fishing-out process operates independently, then the burden of knowledge can be seen as an additional effect constraining the growth rate. Articulated views of why innovation may be getting harder in the growth literature (Kortum, 1997; Segerstrom, 1998) and the innovation literature (*e.g.* Evenson, 1991; Henderson and Cockburn, 1996) have focused on a fishing-out idea. This paper offers the burden of knowledge as a mechanism that makes innovation harder and acts similarly on the growth rate, thus explaining aggregate data trends in addition to the micro-facts presented in this paper.

4. DISCUSSION

This paper is built on two observations. First, innovators are not born at the frontier of knowledge but must initially undertake significant education. Second, the distance to the frontier may vary across fields and over time. Motivated by these observations, I present six novel facts about innovation in cross-section and time series and a model that ties these facts together.

26. In a model with a time cost for education, an increasing burden of knowledge is also felt through increased educational time, as this reduces the portion of the life cycle left over to actively pursue innovations. Jones (2005*b*) considers this more general model.

The burden of knowledge mechanism can further inform several related facts in existing literature. First, consider the age at first invention. Age at first invention is an outcome-based measure intended to delineate the pre-innovation and innovation phases in the life cycle. Alternatively, one might consult an institutionally based measure, such as the age at highest degree. Existing evidence based on doctoral age also suggests an aging phenomenon. Doctoral age rose generally across all major fields from 1967 to 1986, with the increase explained by longer periods in the doctoral programme (National Research Council, 1990). The duration of doctorates as well as the frequency of post-doctorates has been rising across the life sciences since the 1960's (Tilghman, 1998). Nobel Prize winners also show a substantially increasing age at doctorate (Jones, 2005*a*), as shown in Table 1.

The rise in teamwork also generalizes outside patenting institutions, with similarly broad trends reported in academic research. Increasing co-authorship in journal articles is found in virtually all fields of science, engineering, and the social sciences since the 1950's (Wuchty, Jones and Uzzi, 2007). Studies of narrower samples of research fields (*e.g.* Zuckerman and Merton, 1973) suggest that co-authorship has increased steadily since the early 20th century.

The model also provides an explicit analysis of growth, allowing it to inform aggregate facts. In particular, an increasing burden of knowledge can explain why rapid growth in the number of R&D workers and dollars in the 20th century is not associated with increased TFP growth rates or patenting rates (Machlup, 1962; Evenson, 1984; Kortum, 1993; Jones, 1995*b*). The model thus provides a novel solution for the absence of "scale effects", a much debated subject in economic growth. At the same time, the model's analysis of growth is inclusive, incorporating existing mechanisms in the literature regarding innovation exhaustion (fishing-out stories), increasing innovation potential, and market size effects. We see explicitly that the burden of knowledge parameters enter the steady-state growth rate equation much as the parameter capturing any fishing-out effect. Therefore, from a growth perspective, one may view the burden of knowledge mechanism as a micro-foundation for fishing-out-type effects, or, if one imagines a fishing-out process that operates independently, then we can conceive of the burden of knowledge as an additional effect constraining the growth rate.

The model operates with several simplifications to focus on the central mechanisms, but several generalizations are possible. For example, we focus on educational outlays rather than educational duration *per se*; however, educational duration can be incorporated explicitly in a more complex model, and the predictions for innovator behaviour remain the same.[27] The model also places the burden of knowledge mechanism in the rate of idea production and assumes that the size of ideas is fixed. More generally, one may imagine that the burden of knowledge is felt on the size of ideas rather than their rate, or on both dimensions. This generalization is straightforward (Jones, 2005*b*) with no effect on the main propositions and corollaries.

In all, the micro-evidence presented in this paper, together with other available micro-evidence and the aggregate data trends cited previously, suggest general and multi-dimensional patterns that may collectively be understood from the knowledge burden perspective. While any individual piece of evidence can be explained by other means, the burden of knowledge knits together a range of evidence within a single framework. Motivated by the burden of knowledge concept, we are led to a set of striking facts, suggesting large changes in the organization of innovative activity and providing a novel explanation for the absence of scale effects in growth. Moreover, the micro-evidence suggests that the burden of knowledge is increasing. Note that, in general, a combination of increasing specialization and increasing educational attainment is

27. Including time costs of education not only produces the same micro-econometric predictions but also introduces a second dimension through which the burden of knowledge influences growth. As equilibrium educational duration increases along the growth path, the portion of an innovator's life cycle devoted to innovation declines, further restricting the growth rate. This is shown formally in Jones (2005*b*).

difficult to reconcile without appealing to a greater knowledge burden. If the distance to the frontier were not increasing, then increasing education should be associated with broader individual knowledge, not narrowing expertise.

If a rising burden of knowledge is an inevitable by-product of technological progress, then ever increasing effort may be needed to sustain long-run growth. However, two kinds of escapes are worth noting. First, if technological opportunities rise sufficiently rapidly, then the output of innovators may become sufficient, despite a rising educational burden, to sustain growth without increasing effort. While the 20th century's aggregate data patterns—rapidly increasing R&D effort but flat TFP growth—do not suggest a sufficient rise in technological opportunity, there is nothing to say that sufficiently rapid avenues of opportunity may not open in the future.

Second, even if the stock of knowledge accumulates over long periods, some future revolution in science may simplify the knowledge space, causing a fall in the burden of knowledge. Scientific revolutions—Kuhnian "paradigm" shifts (Kuhn, 1962)—might therefore have significant benefits by easing the inter-generational transmission of knowledge. Related to this point, the efficiency of education—the rate at which we transfer knowledge from one generation to the next—becomes a policy parameter with first-order implications for the organization of innovative activity and for growth. Future improvements in knowledge transfer rates could potentially overcome growth in the knowledge stock. While this transfer rate probably faces physiological limits, policy choices in education take on further importance, as policy features from teacher pay to curricular design and the need for a "liberal arts" education all impact the rate at which human capital can be transferred to the young.

APPENDIX A

Proof of Lemma 1. I first show that along a balanced growth path, $L_Y(t) > 0\ \forall t$ and $L_j(t, s_j) > 0\ \forall\ j, s_j, t$. I then show that this implies $U^{\text{R\&D}}(\boldsymbol{\omega}, \tau) = U^{\text{wage}}(\tau)$ for any equilibrium choice $\boldsymbol{\omega}$ and cohort τ.

1. $L_Y(t) > 0\ \forall t$. By contradiction, let $L_Y(t) = 0$ for some t. On a balanced growth path, where labour allocations grow at rate g_L, $L_Y(t) = 0$ for some t implies $L_Y(t) = 0$ for all t. Furthermore, all workers must then be innovators. But these innovators would earn zero income since there is no market for any innovation, with the value of innovations $V(t) = 0$ for all t if $L_Y(t) = 0$ for all t. Therefore, innovators would strictly prefer to be wage workers, earning strictly positive income as defined by equation (4) with $w(t) = X(t - z) > 0$. Hence, by contradiction, $L_Y(t) > 0\ \forall t$.
2. $L_j(t, s_j) > 0\ \forall j, s_j, t$. By contradiction, let $L_j(t, s_j) = 0$ for some j, s_j, and t, which implies that $L_j(t, s_j) = 0$ for all t along a balanced growth path. This cannot hold because being a scarce innovator is too tempting. Recalling that $L_Y(t) > 0$ in any equilibrium, so there is a market for innovations, from the objective function (12), the choice $j, s_j, b_j = 1$ would produce unbounded income. This follows because the choice $b_j = 1$ makes ideas implementable and yet $L_j(t, s_j) = 0$, which makes the rate of idea production unbounded. Hence, a wage worker, who earns bounded income by equation (11) and Assumption 3, would prefer to be such a scarce innovator. Hence, by contradiction, $L_j(t, s_j) > 0\ \forall\ j, s_j, t$.
3. Along a balanced growth path, labour allocations grow at rate g_L. Hence, $L_Y(t) > 0$ implies that individuals in every cohort choose to become wage workers. Meanwhile, $L_j(t, s_j) > 0$ implies that individuals in every cohort choose to be innovators. By the equilibrium conditions of Section 3.5, each choice is weakly preferred to any other and therefore:
$$U^{\text{R\&D}}(\boldsymbol{\omega}, \tau) = U^{\text{wage}}(\tau),$$
for any equilibrium choice $\boldsymbol{\omega}$ and cohort τ. ||

Proof of Proposition 1, *part* (i). As shown in the proof of Lemma 1, $L_j(t, s_j) > 0\ \forall j, s_j, t$. Morever, $b_j > 0$ since otherwise individuals would earn zero income and could do better by choosing to be a production worker, who earns strictly positive income as shown in the proof of Lemma 1. Since individuals exist at every point s_j and $b_j > 0$ in any equilibrium, expertise exists at every point on the circle and all ideas are implementable. Thus, $I_j(t) = 1\ \forall j, t$. ||

Proof of Proposition 1*, part* (ii). By contradiction, imagine that:

$$L_j(t,s') > L_j(t,s''),$$

for two points s' and s'' and some time t. Along a balanced growth path, $L_j(t,s)$ must grow at the population growth rate for any s, which implies that the ratio $L_j(t,s')/L_j(t,s'')$ is constant, and therefore, $L_j(t,s') > L_j(t,s'')$ for all t if $L_j(t,s') > L_j(t,s'')$ for some t. But then anyone located at s' would be strictly better off if they had chosen s''. In particular, the objective function (12) implies that $U^{\text{R\&D}}(j,s',b,\tau) < U^{\text{R\&D}}(j,s'',b,\tau)$ for a person born at time τ. (Given your choice of b, you would always prefer to be at the less crowded location to avoid congestion.) Thus, s' would never be chosen in equilibrium. Therefore, there can be no points on the circle where the mass is less than any other point along a balanced growth path. ||

Proof of Proposition 1*, part* (iii). Differentiating the innovator's objective function, (12), with respect to b_j, produces the first-order condition:

$$\frac{\beta}{b_j}\int_{\tau}^{\infty}\hat{A}_j X(t)^{\chi} L_j^*(t)^{-\sigma} b_j^{\beta} V(t) e^{-\phi(t-\tau)}dt = \frac{\varepsilon}{b_j}(b_j D_j(\tau))^{\varepsilon}, \tag{18}$$

where we have used Proposition 1, parts (i) and (ii), to simplify the objective function; namely, $I_j(t) = 1\ \forall\ j,t$ and $L_j(t,s)$ is invariant with s in equilibrium and written $L_j^*(t)$.

This first-order condition is directly rewritten as $\beta\int_{\tau}^{\infty} \upsilon(\boldsymbol{\omega},t)e^{-\phi(t-\tau)}dt = \varepsilon E(\boldsymbol{\omega},\tau)$. Noting from equation (5) that $\int_{\tau}^{\infty} \upsilon(\boldsymbol{\omega},t)e^{-\phi(t-\tau)}dt = U^{\text{R\&D}}(\boldsymbol{\omega},\tau) + E(\boldsymbol{\omega},\tau)$, the first-order condition is equivalently:

$$E(\boldsymbol{\omega},\tau)/U^{\text{R\&D}}(\boldsymbol{\omega},\tau) = \beta/(\varepsilon-\beta), \tag{19}$$

which is a constant.

The second-order condition takes the form $\partial^2 U^{\text{R\&D}}/\partial b_j^2 = \beta(\beta-1)\ (U^{\text{R\&D}}(\boldsymbol{\omega},\tau)\ + E(\boldsymbol{\omega},\tau))/b_j^2 - \varepsilon(\varepsilon-1)$ $E(\boldsymbol{\omega},\tau)/b_j^2$. The first-order condition thus defines a maximum under Assumption 1:

$$\beta < \varepsilon,$$

which guarantees $\partial^2 U^{\text{R\&D}}/\partial b_j^2 < 0$ where the first-order condition (19) holds.

Note that we may also write the optimal choice b_j as follows. Using equation (19), the definitions of $E(\boldsymbol{\omega},\tau)$ and $D_j(\tau)$ in equations (6) and (7), and the property that $U^{\text{R\&D}}(\boldsymbol{\omega},\tau) = U^{\text{wage}}(\tau)$ in equilibrium, where $U^{\text{wage}}(\tau) = X(\tau)e^{-gz}/(\phi-g)$, we have:

$$b_j^*(\tau) = \frac{X(\tau)^{1/\varepsilon-\delta}}{\hat{D}_j}\left[\left(\frac{\beta}{\varepsilon-\beta}\right)\left(\frac{e^{-gz}}{\phi-g}\right)\right]^{1/\varepsilon}, \tag{20}$$

where we write the optimal choice of b_j as $b_j^*(\tau)$ to clarify that in equilibrium, breadth choices depend on the application area j and cohort of birth τ. Note that $\beta < \varepsilon$ follows from Assumption 1 and $\phi > g$ by Assumption 3, so that $b_j^*(\tau)$ is strictly positive. It is also unique given area of application j and cohort τ. ||

Proof of Corollary 1*, part* (i) *(total education, cross-section).* By Lemma 1, income arbitrage implies $U^{\text{R\&D}}(\boldsymbol{\omega},\tau) = U^{\text{R\&D}}(\boldsymbol{\omega}',\tau)$ for any equilibrium choices $\boldsymbol{\omega}$, $\boldsymbol{\omega}'$ within a given cohort τ. Since $E(\boldsymbol{\omega},\tau)/U^{\text{R\&D}}(\boldsymbol{\omega},\tau)$ is a constant (Proposition 1, part (iii)), $E(\boldsymbol{\omega},\tau) = E(\boldsymbol{\omega}',\tau)$ for any equilibrium choices $\boldsymbol{\omega}$, $\boldsymbol{\omega}'$ within a cohort τ. ||

Proof of Corollary 1*, part* (ii) *(total education, time series).* By Proposition 1, part (iii), $E(\boldsymbol{\omega},\tau)/U^{\text{R\&D}}(\boldsymbol{\omega},\tau)$ is constant. Hence, $E(\boldsymbol{\omega},\tau)$ grows at the same rate as $U^{\text{R\&D}}(\boldsymbol{\omega},\tau)$ across successive cohorts. From equation (11), $g_U = g$. Therefore, $g_E = g$. ||

Proof of Corollary 2*, part* (i) *(breadth of expertise, cross-section).* From equation (6), $b_j = E(\boldsymbol{\omega},\tau)^{1/\varepsilon}/D_j(\tau)$. By Corollary 1, part (i), all innovators in cohort τ have identical $E(\boldsymbol{\omega},\tau) = E^*(\tau)$ regardless of their equilibrium choice of $\boldsymbol{\omega}$. Hence, equilibrium choices of b_j differ only by cohort τ and area of application j. Denoting equilibrium breadth of expertise as $b_j^*(\tau)$, we see that $b_j^*(\tau)/b_{j'}^*(\tau) = D_{j'}(\tau)/D_j(\tau)$. From equation (7), $D_j(t) = \hat{D}_j X(t)^{\delta}$, and hence, $b_j^*(\tau)/b_{j'}^*(\tau) = \hat{D}_{j'}/\hat{D}_j$.[28] ||

28. This result may also be demonstrated directly using the expression for $b_j^*(\tau)$ in equation (20) previously.

Proof of Corollary 2, *part* (ii) *(breadth of expertise, time series).* Taking logs and differentiating with respect to τ, it follows from equation (13) that $g_{b_j^*} = \frac{1}{\varepsilon} g_E - g_{D_j}$. From Corollary 1, part (ii), $g_E = g$, and from equation (7), $g_{D_j} = \delta g$. Therefore, $g_{b_j^*} = (\frac{1}{\varepsilon} - \delta) g$, so that $g_{b_j^*} < 0$ iff $\delta > 1/\varepsilon$.[29] ‖

Proof of Corollary 3, *part* (i) *(teamwork, cross-section).* From equation (14), $\text{team}_j(\tau) = \lceil 1/b_j^*(\tau) \rceil$ if $b_j^*(\tau) \leq 1$ and $\text{team}_j(\tau) = 1$ otherwise. Hence, $\text{team}_j(\tau) \geq \text{team}_{j'}(\tau)$ iff $b_j^*(\tau) \leq b_{j'}^*(\tau)$. From Corollary 2, part (i), $b_j^*(\tau) \leq b_{j'}^*(\tau)$ iff $\hat{D}_j \geq \hat{D}_{j'}$. Therefore, $\text{team}_j(\tau) \geq \text{team}_{j'}(\tau)$ iff $\hat{D}_j \geq \hat{D}_{j'}$. ‖

Proof of Corollary 3, *part* (ii) *(teamwork, time series).* From equation (14), $\text{team}_j(\tau) = \lceil 1/b_j^*(\tau) \rceil$ if $b_j^*(\tau) \leq 1$ and $\text{team}_j(\tau) = 1$ otherwise. Hence, $\text{team}_j(\tau) \geq \text{team}_j(\tau')$ iff $b_j^*(\tau) \leq b_j^*(\tau')$. From Corollary 2, part (ii), $b_j^*(\tau)$ is falling across cohorts iff $\delta > 1/\varepsilon$. Hence, $\text{team}_j(\tau) \geq \text{team}_j(\tau')$ for any $\tau > \tau'$ iff $\delta > 1/\varepsilon$. ‖

Proof of Corollary 4 (labour allocation). Consider the labour allocation across application areas within a given cohort of researchers. Recalling from Proposition 1, part (iii), that $E(\boldsymbol{\omega}, \tau)/U^{\text{R\&D}}(\boldsymbol{\omega}, \tau) = \frac{\beta}{\varepsilon - \beta}$, we can use the definition of $U^{\text{R\&D}}(\boldsymbol{\omega}, \tau)$ in equation (5) to write $U^{\text{R\&D}}(\boldsymbol{\omega}, \tau) = \frac{\varepsilon - \beta}{\varepsilon} \int_\tau^\infty \upsilon(\boldsymbol{\omega}, t) e^{-\phi(t-\tau)} dt$. Rewriting $\upsilon(\boldsymbol{\omega}, t)$ using the definitions in equations (10) and (9) and the equilibrium properties of Proposition 1 we have:

$$U^{\text{R\&D}}(\boldsymbol{\omega}, \tau) = \frac{\varepsilon - \beta}{\varepsilon} \hat{A}_j X(\tau)^\chi b_j^*(\tau)^\beta V(\tau) L_j^*(\tau)^{-\sigma} \int_\tau^\infty e^{\chi g(t-\tau)} e^{(1-\sigma) g_L (t-\tau)} e^{-\phi(t-\tau)} dt, \tag{21}$$

along a balanced growth path, so that $U^{\text{R\&D}}(\boldsymbol{\omega}, \tau)$ for a given cohort varies across applications areas only due to differences in $L_j^*(\tau)$, $\hat{A}_j$, and $b_j^*(\tau)$. Given that $b_j^*(\tau)/b_{j'}^*(\tau) = \hat{D}_{j'}/\hat{D}_j$ (Corollary 2, part (i)), the equilibrium condition $U^{\text{R\&D}}(\boldsymbol{\omega}, \tau) = U^{\text{R\&D}}(\boldsymbol{\omega}', \tau)$ is therefore satisfied by the labour allocation:

$$\frac{L_j^*(\tau)}{L_{j'}^*(\tau)} = \left[\frac{\hat{A}_j}{\hat{A}_{j'}} \left(\frac{\hat{D}_{j'}}{\hat{D}_j} \right)^\beta \right]^{1/\sigma} \qquad \|$$

Proof of Proposition 2 (steady-state growth). To show equation (17), we proceed in three steps. First, it is shown that the rate at which an innovator adds to productivity is the same in equilibrium for all innovators in the same cohort. Second, it is shown that the rate at which innovators add to productivity grows across cohorts at a constant rate. Third, the steady-state growth rate of productivity is determined as the summation of contributions of active innovators.

1. Given common $U^{\text{R\&D}}(\boldsymbol{\omega}, \tau)$ and common $E(\boldsymbol{\omega}, \tau)$ for innovators in the same cohort τ, it follows from the definition of $U^{\text{R\&D}}(\boldsymbol{\omega}, \tau)$ in equation (5) that:

$$\int_\tau^\infty \upsilon(\boldsymbol{\omega}, t) e^{-\phi(t-\tau)} dt = \int_\tau^\infty \upsilon(\boldsymbol{\omega}', t) e^{-\phi(t-\tau)} dt. \tag{22}$$

   Note further that $\upsilon(\boldsymbol{\omega}, t)$ grows with time t at a constant rate independent of ω.[30] Therefore, equation (22) implies $\upsilon(\boldsymbol{\omega}, t) = \upsilon(\boldsymbol{\omega}', t)$ for individuals in the same cohort τ at any point in time t. Writing $\upsilon(\boldsymbol{\omega}, t) = \lambda(\boldsymbol{\omega}, t) V(t)$, where we recall that $\lambda(\boldsymbol{\omega}, t)$ is the rate of idea production, this in turn implies $\lambda(\boldsymbol{\omega}, t) = \lambda(\boldsymbol{\omega}', t)$ for individuals in the same cohort τ at any point in time t.
   Given this result, define the common rate at which innovators in the same cohort add to productivity as $\theta^*(\tau, t)$, where $\theta^*(\tau, t) = \gamma \lambda(\boldsymbol{\omega}, t)$ for any equilibrium choice $\boldsymbol{\omega}$ made at time of birth τ. Noting the definition of $\lambda(\boldsymbol{\omega}, t)$ in equation (9), in equilibrium we have:

$$\theta^*(\tau, t) = \gamma \hat{A}_j X(t)^\chi L_j^*(t)^{-\sigma} b_j^*(\tau)^\beta. \tag{23}$$

2. Inspecting equation (23), the productivity contributions of different cohorts differ only due to their breadth of expertise, $b_j^*(\tau)$. Further, given the results of Corollary 2, part (ii), $b_j^*(\tau)$ grows across cohorts at rate $\left(\frac{1}{\varepsilon} - \delta\right) g$, and hence,

$$\theta^*(\tau, t) = \theta^*(t, t) e^{\beta \left(\frac{1}{\varepsilon} - \delta\right) g(\tau - t)}. \tag{24}$$

29. See prior footnote.

30. In particular, $\upsilon(\boldsymbol{\omega}, t) = \hat{A}_j X(t)^\chi L_j(t, s)^{-\sigma} b_j^\beta V(t)$ is a collection of terms that are either constant given $\boldsymbol{\omega}$ $(\hat{A}_j, b_j)$ or growing at common rates independent of $\boldsymbol{\omega}$ on a balanced growth path $(X(t), L_j(t, s), V(t))$.

3. We can now aggregate the contributions of innovators alive at time t. This is the productivity of each innovator cohort, weighted by the size of that cohort, summed over all living cohorts. Defining $l_R^*(\tau,t)$ as the mass of innovators who were born at time τ who remain alive at time t, we can write the average rate at which all living innovators add to productivity as:

$$\bar{\theta}(t)=\frac{1}{L_R^*(t)}\int_{-\infty}^{t}\theta^*(\tau,t)l_R^*(\tau,t)d\tau.$$

Noting that $L_R^*(t)$ grows at rate g_L and that innovators die at rate ϕ, it is clear that $l_R^*(t,t)=(g_L+\phi)L_R^*(t)$ and $l_R^*(\tau,t)=l_R^*(t,t)e^{(g_L+\phi)(\tau-t)}$.[31] Using these facts and equation (24), we have:

$$\bar{\theta}(t)=\theta^*(t,t)\int_{-\infty}^{t}(g_L+\phi)e^{\beta\left(\frac{1}{\varepsilon}-\delta\right)g(\tau-t)}e^{(g_L+\phi)(\tau-t)}d\tau, \tag{25}$$

where the integral is a finite constant if $\beta\left(\frac{1}{\varepsilon}-\delta\right)g+g_L+\phi>0$, which is guaranteed by Assumption 3, as discussed below.

We are interested in the balanced growth path, where g is a positive constant. Recall from equation (16) that $g=\bar{\theta}(t)L_R^*(t)/X(t)$. Taking logs in equation (16), differentiating with respect to time, and asserting that g is a constant, we see that the steady-state growth rate is:

$$g=g_{\bar{\theta}(t)}+g_{L_R^*(t)}.$$

Noting from equation (25) that $g_{\bar{\theta}(t)}=g_{\theta^*(t,t)}$ and from equation (23) that $g_{\theta^*(t,t)}=\chi g-\sigma g_L+\beta g_{b_j^*}$, we find that $g=\chi g+(1-\sigma)g_L+\beta\left(\frac{1}{\varepsilon}-\delta\right)g$. Rearranging produces the unique steady-state growth rate, the expression (17), repeated here:

$$g=\frac{1-\sigma}{1-\chi+\beta(\delta-\frac{1}{\varepsilon})}g_L$$

where $\chi-\beta(\delta-\frac{1}{\varepsilon})<1$ (Assumption 2). ‖

Existence and uniqueness of balanced growth path. We now confirm that with the steady-state growth rate g defined in equation (17), there exists a set of labour allocations $\{L_1^*(\tau),\ldots,L_J^*(\tau),L_Y^*(\tau)\}$ each growing at rate g_L that satisfies the equilibrium conditions in Section 3.5; namely, that no individual would strictly prefer a different career choice. Moreover, we demonstrate that the set of labour allocations $\{L_1^*(\tau),\ldots,L_J^*(\tau),L_Y^*(\tau)\}$, like the steady-growth rate above, is unique.[32]

We proceed in three steps. First, I confirm that under Assumption 3, lifetime income from different career choices is finite—that is, the integrals in equations (4) and (5) exist. Second, I demonstrate the existence of a unique set of labour allocations $\{L_1^*(\tau),\ldots,L_J^*(\tau),L_Y^*(\tau)\}$ on a balanced growth path such that in a particular cohort τ no individual would strictly prefer a different career choice. Finally, I confirm that the equilibrium conditions are satisfied in all cohorts along the balanced growth path where the steady-state growth rate g is as given in equation (17).

1. The analysis has assumed that lifetime income, $U^{\text{R\&D}}(\boldsymbol{\omega},\tau)$ and $U^{\text{wage}}(\tau)$, is finite. Having defined g explicitly along the balanced growth path, we can now state these assumptions as explicit parametric conditions. First, finite lifetime income for a wage worker requires $\phi>g$ (see equation (11)). Second, finite income for an innovator requires $\phi>\chi g+(1-\sigma)g_L$ (see equation (21)). With g given by equation (17), these conditions are satisfied by Assumption 3. (Related, we assumed previously that $\phi>\beta(\delta-\frac{1}{\varepsilon})g-g_L$ so that the average productivity of innovators is finite (see equation (25)). It is easy to show that this condition is also satisfied by Assumption 3.
2. I now show that, given a steady-state growth rate g, there exists a unique set of labour allocations $\{L_1^*(\tau),\ldots,L_J^*(\tau),L_Y^*(\tau)\}$ for which individuals born at time τ do not strictly prefer another career choice. Note first that Corollary 4 established labour ratios $L_j^*(\tau)/L_{j'}^*(\tau)$ such that $U^{\text{R\&D}}(\boldsymbol{\omega},\tau)=U^{\text{R\&D}}(\boldsymbol{\omega}',\tau)$ for any two equilibrium choices $\boldsymbol{\omega},\boldsymbol{\omega}'$ in cohort τ. Hence, innovators in the same cohort do not strictly prefer to be another type of innovator. This result depends on the labour ratios $L_j^*(\tau)/L_{j'}^*(\tau)$ and not on the overall scale of research,

31. The latter expression follows from two observations: (1) With death rate ϕ, $l_R^*(\tau,t)=l_R^*(\tau,\tau)e^{-\phi(t-\tau)}$ and (2) With growth rate g_L, $l_R^*(t,t)=l_R^*(\tau,\tau)e^{g_L(t-\tau)}$. Hence, $l_R^*(\tau,t)=l_R^*(t,t)e^{(g_L+\phi)(\tau-t)}$.

32. That is, the measures of workers are unique; the equilibrium does not uniquely assign particular individuals to particular careers.

$L_R^*(\tau)$. We can then set $L_R^*(\tau)$, and hence, $L_Y^*(\tau) = L(\tau) - L_R^*(\tau)$, at some unique value to ensure that neither wage workers nor innovators strictly prefer a different career. This is clear because innovator income strictly, continuously decreases over the entire positive real line as $L_R^*(\tau)$ increases while wage worker income is a finite constant.[33] This unique $L_R^*(\tau)$ pins down $L_j^*(\tau)$ for all j and also pins down $L_Y^*(t) = L(t) - L_R^*(t)$. Hence, there exists a unique set of equilibrium labour allocations $\{L_1^*(\tau), \ldots, L_J^*(\tau), L_Y^*(\tau)\}$ for which no individual born at time τ would prefer a different career choice.

3. We have established the existence of a unique set of labour allocations such that $U^{\text{R\&D}}(\boldsymbol{\omega}, \tau) = U^{\text{wage}}(\tau)$ for any equilibrium choice $\boldsymbol{\omega}$ in a particular cohort τ. From equation (11), $U^{\text{wage}}(\tau)$ grows across cohorts at steady-state rate g. Hence, equilibrium will be satisfied in all cohorts if $U^{\text{R\&D}}(\boldsymbol{\omega}, \tau)$ also grows at rate g across cohorts. Taking logs of equation (21), differentiating with respect to τ, and setting equal to g gives $g = \chi g + \beta(\frac{1}{\varepsilon} - \delta)g + g_L - \sigma g_L$. Rearranging this expression reproduces equation (17). Hence, the balanced growth path satisfies the conditions for equilibrium. The steady-state growth rate is as given in equation (17), with the labour allocations $\{L_1^*(t), \ldots, L_J^*(t), L_Y^*(t)\}$ each growing at rate g_L.[34]

APPENDIX B. DATA APPENDIX

The reader is referred to Hall *et al.* (2001) for a detailed discussion of their patent data set. This appendix focuses on the age information collected to augment the data of Hall *et al.*

Age data were collected using the Web site http://www.AnyBirthday.com, which requires a name and zip code to produce a match. As shown in Table B.1, 30% of U.S. inventors listed a zip code on at least one of their patent applications, and of these inventors, http://www.AnyBirthday.com produced a birth date in 25% of the cases. While the number of observations produced by http://www.AnyBirthday.com is large, it represents only 7·5% of U.S. inventors. This appendix explores the causes and implications of this selection. The first question is why zip code information is available for only certain inventors. The second question is why http://www.AnyBirthday.com produces a match only one-quarter of the time. The third question is whether this selection appears to matter.

Table B.2 compares how patent rights are assigned across samples. The table shows clearly that zip code information is virtually always supplied when the inventor has yet to assign the rights; conversely, zip code information is never provided when the rights are already assigned. Patent rights are usually assigned to private corporations (80% of the time) and remain unassigned in the majority of the other cases (17% of the time). An unassigned patent indicates only that the inventor(s) have not yet assigned the patent at the time it is granted. Presumably, innovators who provide zip codes are operating outside binding contracts with corporations, universities, or other agencies that would automatically acquire any patent rights. The zip code subset is therefore not a random sample but is capturing a distinct subset of innovators

33. This can be seen formally as follows. Use equation (15) to write $L_k^*(t) = \left[\frac{\hat{A}_k}{\hat{A}_j}\left(\frac{\hat{D}_j}{\hat{D}_k}\right)^{\beta}\right]^{1/\sigma} L_j^*(t)$. Noting that $L_R^*(t) = \sum_{j=1}^{J} L_j^*(t)$, we can sum across areas of application to write $L_R^*(t) = L_j^*(t)\sum_{k=1}^{J}\left[\frac{\hat{A}_k}{\hat{A}_j}\left(\frac{\hat{D}_j}{\hat{D}_k}\right)^{\beta}\right]^{1/\sigma}$, or $L_j^*(\tau) = cL_R^*(\tau)$ where $c = \left(\sum_{k=1}^{J}\left[\frac{\hat{A}_k}{\hat{A}_j}\left(\frac{\hat{D}_j}{\hat{D}_k}\right)^{\beta}\right]^{1/\sigma}\right)^{-1}$ is a constant. Substituting $L_j^*(\tau) = cL_R^*(\tau)$ into equation (21), we see that $U^{\text{R\&D}}(\boldsymbol{\omega}, \tau)$ is continuous and monotonically decreasing in $L_R^*(\tau)$. This follows because increasing $L_R^*(\tau)$ (a) increases crowding ($L_j^*(\tau)^{-\sigma}$ falls) and (b) reduces the market size for innovations ($L_Y^*(t)$ falls and hence $V(t)$ falls), both of which reduce $U^{\text{R\&D}}(\boldsymbol{\omega}, \tau)$. Moreover, $\lim_{L_R^*(t)\to 0^+} U^{\text{R\&D}}(\boldsymbol{\omega}, \tau) = \infty$ and $\lim_{L_R^*(t)\to L(t)^-} U^{\text{R\&D}}(\boldsymbol{\omega}, \tau) = 0$. Hence, we can pick some $L_R^*(\tau) \in [0, L(\tau)]$ such that $U^{\text{R\&D}}(\boldsymbol{\omega}, \tau)$ is any positive value. Meanwhile, $U^{\text{wage}}(\tau)$ is a finite, strictly positive constant. Hence, there exists some unique scale of research activity, $L_R^*(\tau)$, such that $U^{\text{R\&D}}(\boldsymbol{\omega}, \tau) = U^{\text{wage}}(\tau)$ in the cohort born at time τ.

34. Note that the growth rate, g, is determined two alternative ways. The first approach considers the steady-state growth rate as the summation of contributions of innovators. The second approach considers the steady-state growth rate that guarantees the equilibrium condition $U^{\text{R\&D}}(\boldsymbol{\omega}, \tau) = U^{\text{wage}}(\tau)$ will hold across cohorts. It is instructive to clarify why these two methods produce the same result. First, innovators' total productivity contributions follow the evolution of $\bar{\theta}(t)$, their average rate of productivity contributions per innovator, and $L_R^*(t)$, the scale of research effort. Second, innovator income across cohorts follows the evolution of $\theta(t,t)$ across cohorts and $L_Y^*(t)$, the scale of the market. On the balanced growth path, $L_R^*(t)$ and $L_Y^*(t)$ grow at the same rate. Meanwhile, as shown previously, the evolution of $\bar{\theta}(t)$ with time and the evolution of $\theta(t,t)$ from one cohort to the next are also equivalent. Hence, calculating growth from either perspective produces the same growth rate along a balanced growth path.

TABLE B.1

Number of observations at each stage of selection

	Number of observations	Percentage of row (3)	Percentage of row (4)	Percentage of row above
(1) Patents granted	2,139,313			
(2) Inventors worldwide	4,301,229			
(3) Unique inventors worldwide	1,411,842			
(4) Unique inventors with U.S. address	752,163	53·3		53·3
(5) Unique inventors with U.S. address, zip zode	224,152	15·9	29·8	29·8
(6) Unique inventors with U.S. address, zip code, unique match from http://www.AnyBirthday.com	56,281	4·0	7·5	25·1

Notes: (i) Observation counts consider the 1975–1999 period. (ii) A "unique inventor" is defined by having same first name, last name, and middle initial.

TABLE B.2

The assignment of patent rights

					Birth data	
Assignment status	All patents	U.S. patents	U.S. patents no zip code	U.S. patents zip code	Direct match	Other patents
Unassigned (%)	17·2	22·4	0·4	98·3	97·9	26·6
U.S. non-government organization (%)	43·9	72·9	94·1	0·0	0·0	65·7
Non-U.S. non-government organization (%)	36·2	1·1	1·4	0·0	0·0	3·4
Other assignment (%)	2·7	3·5	4·1	1·7	2·1	4·4

Notes: (i) The first column considers all patent observations in the 1975–1999 period (2·1 million observations). (ii) U.S. patents are those for which first inventor listed with the patent has a U.S. address. (iii) The birth data columns consider those U.S. patents with zip code information for which http://www.AnyBirthday.com produced a birth date. The first birth data column considers the specific patents on which http://www.AnyBirthday.com was able to match. The last column considers all other patents by that innovator, identifying the innovator by last name, first name, and middle initial. (iv) Unassigned patents are those for which the patent rights were still held by the original inventor(s) at the time the patent was granted; these patents may or may not have been assigned after the grant date. (v) Non-government organizations are mainly corporations but also include universities. (vi) Other assignment includes assignments to: (a) U.S. individuals, (b) non-U.S. individuals, (c) the U.S. government, and (d) non-U.S. governments.

who, at least at one point, were operating independently. Despite this distinction, this subset may not be substantially different from other innovators: the last column of Table B.2 indicates that when examining the other patents produced by these innovators, they have a similar propensity to assign them to corporations as the U.S. population average.

The nature of the selection introduced by http://www.AnyBirthday.com is more difficult to identify. The Web site reports a database of 135 million individuals and reports to have built its database using "public records". Access to public records is a contentious legal issue.[35] Public disclosure of personal information is proscribed at the federal level by the Freedom of Information Act and Privacy Act of 1974. At the state and local level, however, rules vary. Birth date and address information are both available through motor vehicle departments and their electronic databases are likely to be the main source of http://www.AnyBirthday.com records.[36] The availability of birth date information is therefore very likely to be related to local institutional rules regarding motor vehicle departments. Geography thus will influence

35. Repeated requests to http://www.AnyBirthday.com to define their sources more explicitly have yet to produce a response.

36. A federal law, the Driver's Privacy Protection Act of 1994, was introduced to give individuals increased privacy. The law requires motor vehicle departments to receive explicit prior consent from an individual before disclosing their personal information. However, the law makes an exception for cases where motor vehicles departments provide information to survey and marketing organizations. In that case, individual's consent is assumed unless the individual has opted-out on their own initiative. See Gellman (1995) for an in-depth discussion of the laws and legal history surrounding public records.

TABLE B.3

Inventors per patent, mean differences between samples

	Dependent variable: inventors per patent				
	(1)	(2)	(3)	(4)	(5)
U.S. address dummy	−0·315	−0·339	−0·300	−0·124	−0·103
	(0·0020)	(0·0020)	(0·0020)	(0·0049)	(0·0048)
U.S. address and zip code dummy	−0·786	−0·670	−0·769	−0·155	−0·176
	(0·0033)	(0·0033)	(0·0032)	(0·0069)	(0·0066)
U.S. address, zip code, and http://www.AnyBirthday.com direct match dummy	0·237	0·246	0·212	0·243	0·228
	(0·0068)	(0·0067)	(0·0067)	(0·0067)	(0·0066)
Constant	2·28	2·57	1·96	1·45	1·56
	(0·0014)	(0·0023)	(0·0052)	(0·0042)	(0·0067)
Technological category dummies	No	Yes	No	No	Yes
Grant year dummies	No	No	Yes	No	Yes
Assignee code dummies	No	No	No	Yes	Yes
R^2	0·0555	0·0825	0·0756	0·0757	0·1162

Notes: (i) Regressions consider means in the entire data set (2·1 million patent observations), covering the 1975–1999 time period. S.E. are given in parentheses. (ii) Dummy variables are nested: the second row captures a subset of the first. The third row captures a subset of the second. (iii) Innovators for whom http://www.AnyBirthday.com produces a birth date are often involved with multiple innovations over the 1975–1999 period. The patents used for comparison in this table are those patents for which http://www.AnyBirthday.com produced the direct match. (iv) Regressions with technological category controls are reported using the six-category measure of Hall *et al.* (2001). Results using the 36-category.

the presence of innovators in the age sample, and a further issue in selection may involve the geographic mobility of the innovator, among other factors. The influence of this selection, together with the implications of assignment status, can be assessed by comparing observable means in the population across subsamples.

Table B.3 considers average team size, which is a source of further differences. Patents with provided zip codes have smaller team sizes than the U.S. average; team sizes in the subset of these patents for which the age of one innovator is known are slightly larger but still smaller than the U.S. average. Controlling for other patent observables, in particular the assignment status, reduces the mean differences and brings the age sample quite closely in line with the U.S. mean. (See the last two columns of the table.) Having examined a number of other observables in the data, such as citations received and average tree size, I find that relatively small differences tend to exist in the raw data and that these can be either entirely or largely explained by controlling for assignment status and team size. Most importantly, the age results in the text are all robust to the inclusion of assignment status, team size, and any other available controls.

Finally, examining team size, specialization, and time lag trends in the age subsample, the results are similar in sign and significance as those presented in Section 4. The rate of increase in specialization is larger, and the rate of increase in team size is smaller. The time lag shows no trend. Reexamining trends in the entire data set by assignment status, I find that the team size trend is weaker among the unassigned category, which likely explains the weaker trend in the age subset. Similarly, I find that the specialization trend is stronger among the unassigned category, which likely explains the stronger trend in the age subset.

I conclude therefore that while the age subset is not a random sample of the U.S. innovator population, the differences tend to be explainable with other observables and, on the basis of including such observables in the analysis, the age results appear robust.

Acknowledgements. I thank Ricardo Alonso, Pol Antras, Andrei Bremzen, Esther Duflo, Glenn Ellison, Amy Finkelstein, Simon Johnson, Jin Li, Joel Mokyr, Ben Olken, Michael Piore, and Scott Stern in addition to Maitreesh Ghatak and three anonymous referees for helpful comments. I am especially grateful to Daron Acemoglu, Abhijit Banerjee, and Sendhil Mullainathan for their advice, and Trevor Hallstein for research assistance. The support of the Social Science Research Council's Program in Applied Economics, with funding provided by the John D. and Catherine T. MacArthur Foundation, is gratefully acknowledged.

REFERENCES

AGHION, P. and HOWITT, P. (1992), "A Model of Growth through Creative Destruction", *Econometrica*, **60**, 323–351.

ARROW, K. J. (1962), "Economic Welfare and the Allocation of Resources for Invention", in R. Nelson (ed.) *The Rate and Direction of Inventive Activity: Economic and Social Factors* (Princeton, NJ: Princeton University Press) 609–625.

BARRO, R. and SALI-I-MARTIN, X. (1995), *Economic Growth* (New York, NY: McGraw Hill).
BEN-PORATH, Y. (1967), "The Production of Human Capital and the Life Cycle of Earnings", *Journal of Political Economy*, **75**, 352–365.
BLANCHARD, O. J. (1985), "Debt, Deficits, and Finite Horizons", *Journal of Political Economy*, **93**, 223–247.
EVENSON, R. E. (1984), "International Invention: Implications for Technology Market Analysis", in Zvi Griliches (ed.) *R&D, Patents, and Productivity* (Chicago, IL: University of Chicago Press) 89–123.
EVENSON, R. E. (1991), "Patent Data by Industry: Evidence for Invention Potential Exhaustion?" in *Technology and Productivity: The Challenge for Economic Policy* (Paris: OECD) 233–248.
GELLMAN, R. (1995), "Public Records: Access, Privacy, and Public Policy", *Government Information Quarterly*, **12**, 391–426.
HALL, B. H., JAFFE, A. B. and TRAJTENBERG, M. (2001), "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools" (NBER Working Paper No. 8498).
HART, O. and MOORE, J. (1994), "A Theory of Debt Based on the Inalienability of Human Capital", *Quarterly Journal of Economics*, **109**, 841–879.
HENDERSON, R. and COCKBURN, I. (1996), "Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery", *Rand Journal of Economics*, **27**, 32–59.
JONES, B. F. (2005*a*), "Age and Great Invention", *Review of Economics and Statistics* (forthcoming).
JONES, B. F. (2005*b*), "The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder?" (NBER Working Paper No. 11360).
JONES, B. F. (2008), "The Knowledge Trap: Human Capital and Development, Reconsidered" (NBER Working Paper No. 14138).
JONES, C. I. (1995*a*), "R&D-Based Models of Economic Growth", *Journal of Political Economy*, **103**, 759–784.
JONES, C. I. (1995*b*), "Time Series Tests of Endogenous Growth Models", *Quarterly Journal of Economics*, **110**, 495–525.
KORTUM, S. S. (1993), "Equilibrium R&D and the Decline in the Patent-R&D Ratio: U.S. Evidence", *American Economic Review Papers and Proceedings*, **83**, 450–457.
KORTUM, S. S. (1997), "Research, Patenting, and Technological Change", *Econometrica*, **65**, 1389–1419.
KUHN, T. (1962), *The Structure of Scientific Revolutions* (Chicago, IL: University of Chicago Press).
MACHLUP, F. (1962), *The Production and Distribution of Knowledge in the United States* (Princeton, NJ: Princeton University Press).
MALONE, M. S. (1995), *The Microprocessor: A Biography* (New York, NY: Springer-Verlag).
NATIONAL RESEARCH COUNCIL (1990), *On Time to the Doctorate: A Study of the Lengthening Time to Completion for Doctorates in Science and Engineering* (Washington, DC: National Academy Press).
NORDHAUS, W. D. (1969), "An Economic Theory of Technological Change", *American Economic Review*, **59**, 18–28.
ROMER, P. M. (1990), "Endogenous Technological Change", *Journal of Political Economy*, **98**, S71–S102.
SCOTCHMER, S. (1991), "Standing on the Shoulders of Giants—Cumulative Research and Patent Law", *Journal of Economic Perspectives*, **5**, 29–41.
SEGERSTROM, P. (1998), "Endogenous Growth without Scale Effects", *American Economic Review*, **88**, 1290–1310.
TILGHMAN, S. (1998), *Trends in the Early Careers of Life Sciences* (Washington, DC: National Academy Press).
WEITZMAN, M. L. (1998), "Recombinant Growth", *Quarterly Journal of Economics*, **113**, 331–360.
WUCHTY, S., JONES, B. F. and UZZI, B. (2007), "The Increasing Dominance of Teams in the Production of Knowledge", *Science*, **316**, 1036–1039.
YOUNG, A. (1998), "Growth without Scale Effects", *Journal of Political Economy*, **106**, 41–63.
ZUCKERMAN, H. and MERTON, R. (1973), "Age, Aging, and Age Structure in Science", in R. Merton (ed.) *The Sociology of Science* (Chicago, IL: University of Chicago Press) 497–559.