BIG IDEAS OR SPECIALIZATION: WHY ARE INVENTOR TEAM SIZES INCREASING?

Jared S. Franz, University of Illinois at Chicago

ABSTRACT

This paper investigates the reason for increasing team size in the U.S. patent data. On the one hand, Jones (2005) finds that greater specialization over time by R&D workers increases team size. Alternatively, team size can increase because R&D workers are getting larger ideas over time too. A large idea is an idea that requires a large breadth of expertise to implement. R&D workers who conceive large ideas are connecting distant knowledge from outside their own area of specialization and as a result must form large teams to implement these ideas too. To measure the size of ideas, this paper uses the Olsson model to create technological distance measures which proxy for idea size and finds that technological distances are increasing. Thus, greater specialization only partially explains the increases in team size.

INTRODUCTION

The Jones model of growth focuses on the R&D worker as the unit of analysis where new ideas increase productivity in the economy. One feature of the model is that it predicts increasing team size over time as knowledge matures; and, Jones finds empirical support for this in the U.S. patent data over the period 1975-1999 (2005).

The decision to form a team is straightforward: an R&D worker chooses to form a team if the idea he conceives has a larger breadth of expertise than his own breadth of expertise obtained from his years in school. If it does, the R&D worker forms a team that together covers the necessary breadth of expertise needed to implement the idea. An idea that requires a large breadth of expertise to implement connects distant knowledge relative to the R&D worker's specialty. For the purpose of this paper these ideas are referred to as 'big ideas'.

For example, a big idea might connect knowledge in computer science with knowledge in chemistry. In order for an R&D worker with a specialty in chemistry to implement this idea he must form a team with a computer scientist (at the least). On the other hand, seemingly small ideas may require a team if the degree of specialization is large, for example, a specialist in inorganic chemistry may need to form a team to implement an idea he has in analytical chemistry because he does not have the required expertise.

As Jones notes, one reason that team size could be increasing is that as knowledge becomes more complex individual R&D workers choose a smaller breadth of expertise, i.e., they specialize in narrower areas of knowledge. Hence, they rely on increasing larger teams to implement the ideas they conceive that are outside their increasingly smaller specialization. However, if individual R&D workers are conceiving bigger ideas over time, this 'big ideas' effect would require larger teams too. Overall, the increase in team size documented by Jones could arise from 2 separate (simultaneous) effects: 1) increase in specialization and 2) increase in idea size.

To investigate the second point this paper uses a set-theoretic approach developed by Olsson (2000; 2005) to construct technological distance measures from the U.S. patent data that proxy for idea size. Overall, 3 different technological distance measures are constructed for each utility patent by measuring the distance between the citing patent class and each cited patent class. Thus, a new patent in patent class 71: chemistry fertilizer that makes a citation to a patent in patent class 455: telecommunications has a large technological distance and is more likely to require a team of R&D workers to implement. On the other hand, a new patent in patent class 71: chemistry fertilizer that makes a citation to a patent in the same patent class 71: chemistry fertilizer is more likely to be invented by solo R&D worker since the technological distance is small.

This paper argues that these technological distances are a good proxy for idea size in the Jones model for 2 reasons: 1) technological distance is positively related to team size: larger distances require larger teams, and 2) distances for solo R&D workers are smaller than distances for teams of R&D workers. While this paper presents preliminary results both properties hold on average for technological distance created from U.S. patents in 1975 and 1995.

The time trend of technological distance from 1975 – 1995 provides strong evidence that idea size is increasing over time. This means that a joint explanation for increasing team size arising from specialization and large ideas is likely, i.e., team sizes are increasing because R&D workers are specializing and because R&D workers are getting larger ideas over time. Other findings reveal a specialization paradox concerning solo R&D workers: if increases in the complexity of knowledge were forcing R&D workers to specialize the data should reveal a decrease in the technological distances solo R&D workers implement by themselves. However, this study indicates that solo R&D workers are increasingly implementing ideas with larger technological distances at the same pace as teams of R&D workers (though smaller in absolute terms).

THE JONES MODEL

The following model of economic growth is based on Jones (2005). The distinguishing feature of this model is that it focuses on the negative impact of the burden of knowledge on successive generations of R&D workers. This section outlines the model in general emphasizing the results that pertain to increasing team size.

Consider the typical firm which produces a homogeneous good, $y_j(t) = X_j(t) l_j(t)$, where, $y_j(t)$, is the amount of output produced by firm *j*, and, $X_j(t)$, is the productivity level of firm *j*, and, $l_j(t)$, is the amount of labor hired by firm *j*. The total output for the entire economy is:

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

The variable, $X(t)$, is the leading edge of technology for the entire economy, $L_Y(t)$, is the aggregate productive labor force. The revenues of the firm are completely exhausted in wage payments and royalty payment to production workers and R&D workers respectively, $X(t) l_j(t) = w(t) l_j(t) + r(t) l_j(t)$, where, $w(t)$, is the wage to production workers, and, $r(t)$, is the royalty payment to R&D workers for use of a patented invention. The wage paid to a production worker is:

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

Workers in the model get utility from the present value of expected lifetime non-interest income with a hazard rate of death, ϕ, defined by the following general intertemporal utility function at the time of birth τ:

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

Since the economy has 2 types of workers with differing sources of income the utility function will depend on whether it is income from a production workers or income from a R&D worker. The choice of career is decided at birth so that production workers receive lifetime income from wages,

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

On the other hand, R&D workers receive an expected flow of income, $v_i(t)$, from licensing their invention to the productive sector. The difference between production workers and R&D workers is that R&D workers must pay educational costs, $E_i(t)$, to get to the frontier of knowledge in order to conceive ideas that generate income. Since income to R&D workers depends on their field of specialization, s_i, and the associated educational costs for becoming an expert in that field, utility varies for different R&D workers.

$$U_i^{R\&D}(\tau) = \int_{\tau}^{\infty} v_i(t)\, e^{-\phi(t-\tau)}\, dt - E_i(\tau) \tag{5}$$

For R&D workers, getting to the frontier of knowledge is expensive if the distance to the frontier is large. In addition, if the distance to the frontier is large and the individual decides to study a wide breadth of knowledge then educational costs will be even higher. For example, an individual who decides to double major in chemistry and math will have a higher educational cost than a single major in either assuming the depth of knowledge in each field is the same. Thus the educational costs for individual, *i*, depends on the (chosen) breadth of expertise, b_i, and the depth of knowledge in that specialty at the time of birth, $D(\tau)$. Educational cost are determined as,

$$E_i(\tau) = (b_i D(\tau))^{\varepsilon} \tag{6}$$

The elasticity of educational costs to learning more knowledge is captured by, $\varepsilon > 0$.

Over time the depth of knowledge in any field of specialization, s_i, changes and differs from other fields. For example, in 1950 the depth of knowledge in the field of computers was minimal compared to what it is today. On the other hand, in 2007 the depth of knowledge in computers might be considered deeper than in economics. For now, the model assumes that the depth of knowledge is equal in all fields, and the total depth of knowledge at any moment in time is given by, $D(t)$.

New ideas in all fields of knowledge change the total depth of knowledge, $D(t)$. In some areas, new ideas might completely replace old ideas (reducing the depth); but, in other areas new ideas might be viewed as complementary to old ideas (increasing the depth). Regardless of their impact on depth, new ideas increase productivity, $X(t)$, in the economy with, δ, capturing the direction of the impact on the depth:

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

To summarize, R&D workers must bring themselves to the frontier of knowledge in their chosen specialty before they can create new innovative ideas. Once an R&D worker is educated the decision to form a team depends on their own breadth of expertise and the size of the ideas they conceive.

Ideas, Expertise, and Teams

In general, the decision by R&D workers to pick an area of specialization, s_i, and a breadth of expertise, b_i effects the ideas they can implement. The larger their breadth of expertise the more likely they can implement an idea without needing assistance from anyone. In the limiting case of omnipotence, a solo R&D worker can implement any idea by himself. However, if the R&D worker does not have the breadth of expertise to implement the idea then he forms a team that has the necessary expertise to implement the idea. Specifically, the R&D worker can only implement a new idea if his breadth of expertise, b_i, encompasses the breadth of expertise needed to implement the idea.

Jones assumes that an idea, i_n, arrives randomly with a hazard rate, λ, to the R&D worker with the following 2 properties: 1) a random breath of expertise, k, needed to implement the idea and 2) γ, which reflects the impact of the idea on economic productivity. The random breadth of expertise, k, is drawn from a smooth distribution function, *F*. If the R&D worker has the required expertise, $k \le b_i$, he develops the idea as a solo R&D worker; however, if the R&D worker does not have the required expertise, $k > b_i$, he forms a team of R&D workers to implement the idea.

The model does not assume how R&D workers form new ideas, only that the new idea, i_n, comes from within the R&D worker's breadth of expertise. Thus, someone trained as a chemist does not get new ideas completely within computer science; however, a chemist can get ideas that combine knowledge in chemistry and computer science. This paper assumes that new ideas are formed as a combination of ideas, and where at least one part of the combination is in the R&D worker's specialty. This process can be a deliberate act by the R&D worker or can happen passively in the R&D worker's mind. Thus, when a new idea arrives to the R&D worker the random breadth of expertise, k, is the distance between the knowledge used to create the idea. The solo or team decision still depends the R&D worker's breadth of expertise, b_i, relative to, k.

Regardless of how the idea is implemented (solo or team) once it takes the form of a patent it can be licensed to the productive workers of the economy. The product market of the invention has a size, $M(t)$, determined by the number of productive workers, $L_Y(\tilde{t})$, that the invention can be licensed to for the life of the patent, z.

$$M(t) = \int_t^{t+z} L_Y(\tilde{t})\,d\tilde{t} \tag{8}$$

The lump-sum value of the patent is, $V(t) = \gamma M(t)$, depends on the size of the market and the impact of the idea on economic productivity. The expected flow of income to the R&D worker is, $v = \lambda V$, which can be re-written to be a function of 3 things: the probability of getting an idea, the impact of the idea on productivity, and the total market size for the inventions, $v = \lambda\gamma M(t)$. Furthermore, Jones defines R&D worker's individual productivity as, $\theta_i(t) = \lambda\gamma$, which is the impact the R&D worker's idea will have on the economy and is parameterized in the following way:

$$\theta_i(t) = X(t)^{\chi} L(t, s_i)^{-\sigma} b_i^{\beta} \quad (9)$$

The innovative capacity of an individual R&D worker takes a linear multiplicative form determined by the current level of productivity, $X(t)$; the mass of individuals sharing the R&D worker's specialty, $L(t, s_i)$; and, the R&D workers chosen breadth of expertise, b_i. The parameter, χ, measures the impact of new ideas on individual productivity. The parameter, σ, measures the impact of crowding in one's specialty on individual productivity. Lastly, the parameter, β, measures the impact of specialization on individual productivity.

Combining everything, the expected flow of income to R&D worker *i*,

$$v_i(t) = X(t)^{\chi} L(t, s_i)^{-\sigma} b_i^{\beta} M(t) \quad (10)$$

The equilibrium conditions for solving this problem are explained in Jones (2005) and will not be restated here except to say that R&D worker maximizes the following equation at birth,

$$\max_{s_i, b_i} \int_{\tau}^{\infty} X(t)^{\chi} L(t, s_i)^{-\sigma} b_i^{\beta} M(t) e^{-\phi(t-\tau)}\, dt - (b_i D(\tau))^{\varepsilon} \quad (11)$$

The variable of interest for this paper is the equilibrium choice of the breadth of expertise, b_i^*. More specialization implies that the value of, b_i^*, is getting smaller over time which means larger teams will be needed. Jones shows that R&D worker specialization will occur "if the depth of knowledge, rises relatively quickly given the ease with which knowledge can be learned" (2005), i.e., $\delta > 1/\varepsilon$. In addition, corollary 1 of Jones shows that along a balanced growth path of the economy the following conditions holds,

$$\overline{team}(t) > 0 \;\; iff \;\; \delta > 1/\varepsilon \;\; (specialization)$$

This result which Jones states as "the behavior of the average team size, $\overline{team}(t)$, follows the same condition as specialization...more specialized workers rely on teamwork for the implantation of their ideas" (2005). But, recall that the mechanism that causes team size to increase in the Jones model is the struggle that every R&D worker faces – implementing their ideas. As the value of, b_i^*, is getting smaller the probability of, $k > b_i$, gets larger. However, by focusing exclusively on specialization and without a measure for, k, it is hard to know if this is true. If R&D workers are getting increasingly larger ideas over time then team size will increase and the impact of specialization on team size will be reduced.

In order to determine if this is happening, I propose nesting the Olsson model of technological distance into the Jones model. The purpose of using the Olsson model is to develop theoretical and empirical justification for constructing a proxy for the variable,

k, in the Jones model and thus determine if increasing team sizes observed in the U.S. patent data comes from the impact of larger ideas.

THE OLSSON MODEL

Olsson's model constructs a multidimensional idea space where the distance between ideas can be measured. If the distance between 2 ideas is small then these ideas are close together in idea space and thus are more closely related than if they were far apart. Ideas that have a small technological distance require a smaller breadth of expertise to implement. When ideas arrive randomly to the R&D worker in the Jones model they arrive with a random breadth of expertise, k, that is measured in the Olsson model as a technological distance.

The problem with using the Olsson model is that technological distance is difficult to calculate because it requires the practitioner to score every idea based on the 'dimension of human thinking' (2005) – though theoretically feasible, practically it is impossible. My approach - while keeping within the mathematical framework of the Olsson model - introduces the theoretical concept of classes and uses these classes to construct 3 distance measures (maximum, total, and average). Instead of scoring a new idea based on the 'dimensions of human thinking', this approach assumes that similar ideas will share the same area in idea space.

That is, ideas related to chemistry are close together in idea space and they share the same technological class: chemistry. Zooming out from this technological class would eventually incorporate ideas in chemistry and physics in the same technological class: hard sciences. Zooming out even more would incorporate a wide array of ideas from chemistry to economics. Therefore, measuring the technological distance between 2 ideas in idea space simplifies to determining whether the 2 ideas occupy the same technology class, or not. For example, the technological distance between chemistry and physics is smaller than between chemistry and economics. The next few sections will describe the construction of idea space and technological distance.

Idea Space

*Definition 1: An **idea** is a metaphysical outcome of a thinking process that has been, is, or might be stored by at least one human brain.*

*Definition 2: $\mathcal{I}$ is the infinite, **universal set** of all possible ideas.*

A key assumption of this model is that all human thinking produces ideas, i.e., thinking about lunch is an idea. Or, (you) thinking about a cure for obesity is an idea: most people produce mundane ideas that contribute nothing to society's flow of knowledge.

*Definition 3: I is the Euclidean metric space, or the **idea space**, associated with $\mathcal{I}$ such that $I \subset \Re_{+}^{v}$.*

*Definition 4: The **location** of an idea $m \in \mathcal{I}$ in the idea space I is i_m.*

Olsson notes that the difference between, $\mathcal{I}$, the universal set of ideas, and I, the idea space is that the universal set of ideas is not a metric space, that is, the idea *m* has no dimensionality in $\mathcal{I}$. Defining a metric space, I, gives an idea, $m \in \mathcal{I}$ dimensionality which can be used to compare the idea, *m*, to other ideas in, $\mathcal{I}$. Each idea, $m \in \mathcal{I}$, in the universal set of ideas, maps to a unique location in, I, the idea space. The location of an idea in idea space is defined by the dimensions of human thinking (explained below), and not by what the object looks like. For example, the sound of water (the outcome of a thinking process) is not an object, but is still an idea (by definition 1).

The location of any idea, i_m, in idea space is a *v*-dimensional column vector, where the *v*-dimensions represent the dimensions of human thinking. For example, consider any idea, $i_m \in I$, and assume a world where there are only 3 dimensions of human thinking (*v*=3): complex (c), labor (l), and mechanical (a). Any idea in this simple world has the following column vector representation:

$$i_m = \begin{bmatrix} complex\ (c) \\ labor\ (l) \\ mechanical\ (a) \end{bmatrix}_{v=3}$$

If the idea, i_m, relies heavily on the mechanical dimension of human thinking, then the mechanical coordinate (a) is a large real number. On the other hand, if the idea relies on the complex dimension of human thinking, then the coordinate (c) is a large real number. Likewise, all ideas in this (simple) world are defined using this method, that is, every idea is judged based on the same 3 dimensions of human thinking. The numeric value of each of the 3 coordinates (complex, labor, and mechanical) plots a single point in idea space. With only 3 dimensions of human thinking, the idea space for this world is a 3-dimensional graph with axes representing the complex dimension, labor dimension, and mechanical dimension.

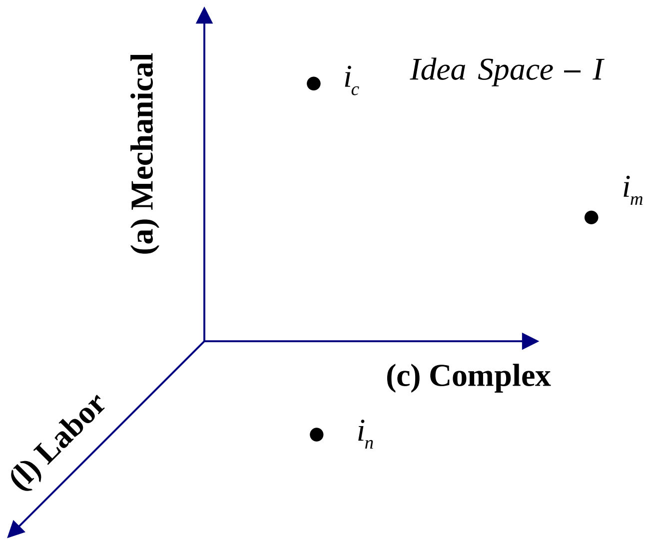


Figure 1. Idea Space for v=3.

The placement of the 3 ideas (i_c, i_m, i_n) in figure 1 gives an indication of what ideas in idea space could look like. The idea, i_c, has a higher numerical value for the mechanical dimension of human thinking than the other two ideas, but has about the same level of complexity as the idea i_n. All ideas, defined by definition 1, can be placed in this simple 3-dimensional space; in this example, the dimensions of human thinking are small and the three dimensions chosen are only illustrative. Alternatively, one could pick electrical, chemical, and medical as the three dimensions in figure 1 and re-classify the three ideas (i_c, i_m, i_n).

*Definition 5: For any 2 ideas $i_l, i_m \in I$ the **distance function** $d(i_l, i_m) \in \Re_+$ is the distance between i_l and i_m, which satisfies the following conditions: (i) $d(i_l, i_m) \geq 0$ and $d(i_l, i_m) = 0$ iff $i_l = i_n$, (ii) $d(i_l, i_m) = d(i_m, i_l)\ \forall\ i_m\ i_l \in I$, and (iii) $d(i_l, i_m) + d(i_m, i_o) \geq d(i_l, i_o)$.*

*Definition 6: Given $\varepsilon > 0$ and an idea i_m in the metric space (I, d), the **ε-neighborhood** of i_m is the set $V_\varepsilon(i_m) = \{i_l \in I : d(i_m, i_l) < \varepsilon\}$.*

*Definition 7: A **class** is a set of similar ideas $V_c(i_q) \in I$ defined by $c \in \Re_+$.*

The distance function is how the breadth of expertise will be measured and the class definition simplifies how the distance between ideas is calculated.

The Knowledge Set

Definition 1 states that the metaphysical outcome of a thinking process is an idea; but, this definition is too broad to be useful, since an R&D worker can fill idea space with thousands of ideas daily. The next set of definitions and assumptions constrains the idea space, I, to ideas that are useful to science at a given moment in time: the knowledge set is the most important subset of idea space.

*Definition 8: The **knowledge set** is* A_t *which has the following characteristics: (i)* $A_t \subset I \subset \Re_+^v$ *,i.e., the knowledge set is a subset of idea space* I *, which is defined in metric space* $\Re_+^v$ *. (ii)* $I = A_t \cup A_t^C$ *, i.e., the knowledge set and its complement is what constitutes idea space. (iii)* A_t *is infinite, closed, and bounded.*

The knowledge set, A_t, is all useful knowledge as of time t, or as Olsson states: "all ideas embraced by science at time t" (2000). Here science refers to the broad European definition of the word meaning systematic inquiry (McCloskey, 2000) and not to the specific Anglo-American definition usually referring only to biology, physics, chemistry, and the other hard-sciences. Accordingly, any field of systematic inquiry contributes knowledge to the knowledge set: ideas in political science, literature, and economics are in the knowledge set, A_t. Furthermore, research and development at firms where systematic inquiry is used to improve a consumer product, make a process improvement, or streamline supply chains is also considered science.

Incremental Innovation in the Knowledge Set

In general, the knowledge set in the Olsson model expands over time from R&D workers getting ideas in the Jones model. This knowledge creation in the Olsson model is modeled explicitly as incremental innovation where new ideas are a convex combination of old ideas on the boundary of the knowledge set (the knowledge frontier). The distance between the connected ideas is measured by the distance function, $d(\cdot)$, in idea space which is the breadth of expertise, k, in the Jones model.

*Assumption 1: An **incremental innovation** is a linear, binary combination of two technologically close ideas* $i_l, i_c \in bdy(A_t)$ *such that (i)* $d(i_l, i_c) \leq \bar{d}$ *, and (ii) that the newly formed idea* $\omega i_l + (1-\omega) i_c = i_n \notin A_t$ *where* $\omega \in (0,1)$*.*

Once a new idea is created the region formed by the boundary of the knowledge set and the idea will be exploited quickly thereafter (Olsson, 2000), i.e., the entire interior region will become part of the knowledge set depicted in figure 2. The reason for this might be imitators copying the idea with minor changes to make it different. This behavior occurs frequently in the patent literature with attempts to patent around an existing patent. A long distance connection may cause a flurry of subsequent ideas that slightly modify the original long distance idea; the longer the distance between any 2 ideas the larger the space that is carved out of the region.

There is an upper limit to the distance between ideas that an R&D worker can connect in assumption 1. Olsson notes the following: "the upper bound of, $\bar{d}$, might be thought of as the longest distance between two ideas that an R&D worker is capable of crossing for a combinational attempt" (2005). The distance that a single R&D worker can traverse may depend on the education, the intelligence, and the creativity of the R&D worker engaged in systematic inquiry. I equate the upper bound, $\bar{d}$, as the limit of individual bounded rationality, which constrains the amount of information individuals can process and limits the distance that individuals by themselves can traverse.

As noted earlier, the conception of an idea, i_n, in assumption 1 is different from the implementation of the idea. The implementation (solo or team) of the idea, i_n, depends on the technological distance between the ideas used to create it, $d(i_l, i_c)$, relative to the R&D workers own breadth of expertise, b_i. R&D workers who form teams can implement ideas that have larger technological distances relative to their own breadth of expertise. However, if the technological distance is small (within his own expertise) then the idea is implemented without a team.

A pictorial representation of both situations is presented in figure 2. Here the boundary of the knowledge set, $bdy(A_t)$, represents the frontier of knowledge for particular knowledge discipline. The circle created by the radius, c, depicts the collection (or class) of ideas that an R&D worker has learned and is analogous to, b_i, the R&D worker's chose breadth of expertise. The technological distance, $d(i_l, i_q)$, depicts the linear distance between 2 ideas that were convexly combined to form new knowledge, i_n, according to assumption 1. The technological distance between the 2 ideas, $d(i_l, i_q)$, is analogous to, k, the breadth of expertise required to implement the idea. If the new idea, i_n, comes from knowledge within the R&D worker's breath of expertise, b_i, (represented by the circle) then it can be implemented as a solo invention. On the other hand, if the R&D worker conceives an idea the combines more distant knowledge, $d(i_l, i_m)$, in figure 2 which extends outside the R&D worker's breadth of expertise, b_i, then it requires a team to implement.

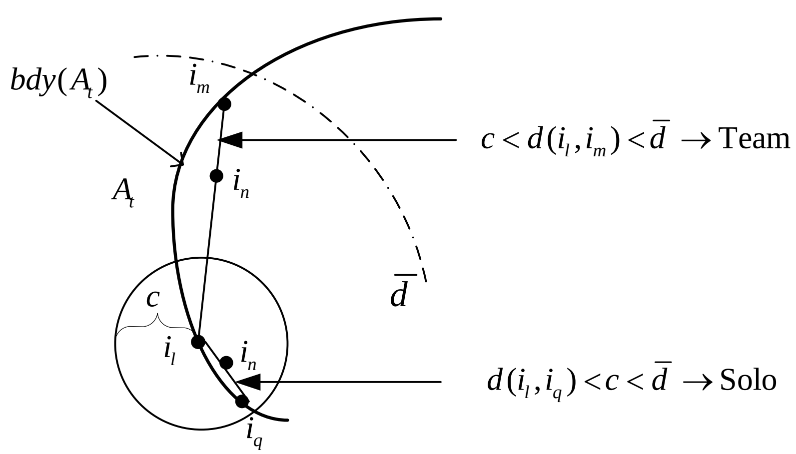


Figure 2. Implementing Inventor Ideas.

Specialization over time works by reducing the size of, c, in figure 2; thus, making it harder to implement ideas alone. On the other hand, increasing idea size works by increasing the technological distances, $d(\cdot)$, that R&D workers are conceiving.

Previously measuring the distance between ideas, $d(\cdot)$, gave an indicator of similarity; however, in practice this distance measure is nearly impossible to implement. This stems from the complexity of a large *v*-dimensional idea space and assigning ideas coordinate values based on each dimension of human thinking. An easier (though less precise) way is to use the definition of classes. For example, starting from an arbitrary idea, i_l, in idea space and defining a small value, $c \in \Re_+$, creates a class of nearly exact ideas, $V_c(i_l)$. Increasing the size of c creates a larger class of similar ideas around the idea, i_l. A still larger value of c includes ideas that are very different from the original idea, i_l; but, where all of these ideas are still related.

But, instead of just increasing the size of c, use breakpoints that define specific larger and larger classes as depicted in figure 3 but with patent notation. For example, the first breakpoint, c_1, defines a radius that captures all ideas that are nearly exact into one class, $V_{c_1}(i_l)$. The second breakpoint, c_2, defines a slightly larger radius that captures all ideas that are similar though not exact into one class, $V_{c_2}(i_l)$. From this point of view, the distance between ideas depends on whether an idea is within a particular radii of another idea: an idea that is located within the c_1 radius is more closely related to, i_l, than an idea within the c_3 radius. Furthermore, if the distance between, i_l, and each breakpoint radius is known then this would give a general measure of how far ideas are from each other. Having a large number of breakpoints adds precision to this technique without resorting to labeling each idea based on the dimensions of human thinking.

Admittedly, the cost of this technique is precision: determining the distance between ideas *within* the same class is difficult. On the other hand, this technique provides a good way to measure ideas that are far apart. Going forward calculating the distance between ideas in idea space will use the distance technique described above (and not by the dimensions of human thinking described by Olsson).

Since patents are a subset of the ideas conceived and implemented during a specific time period this paper constructs 3 technological distance measures for over 2 million U.S. patents granted between 1975 and 1999 based on the patent classes and aggregation method suggested by Jaffe and Trajtenberg (2002). In addition, for each distance constructed the associated team size was calculated based on the number of R&D workers on the patent.

METHODOLOGY

The previous sections outlined the theory of the Olsson model and introduced a method for calculating distance within the Olsson model of ideas. This section applies the distance technique to the U.S. patent data, meaning that R&D workers engage in systematic inquiry and convexly combine old knowledge (cited patents) to produce new ideas (citing patents). In figure 3, the idea notation, $i_m \in A_t$, is switched to patent notation, $p_m \in A_t$.

NBER U.S. Patent Data 1975 – 1999

To my knowledge there are 2 locations to download this data: the National Bureau of Economic Research (NBER) and the book Patent, Citations, & Innovations by Jaffe and Trajtenberg. Each source has consolidated and made available 5 large patent files in 3 formats. The NBER source is on the web at http://www.nber.org/patents/; the file formats are (1) SAS .tpt format and (2) ASCII II .csv format. The second source has the data stored in 1 additional format available with the CD-ROM that comes with the book: (3) dBase .dbf format. The original data for both sources is the United States Patent and Trademark Office (USPTO). All database work for this paper was done in SQL using the ASCII data from the NBER.

The 5 patent files are: (1) Pat63_99, (2) Cite75_99, (3) Coname, (4) Match, and (5) Inventor. The main file Pat63_99 contains detailed patent level information for all utility patents (nearly 3 million) granted between 1963 and 1999. The file has 23 variables, the first 10 variables are the original variables downloaded directly from the USPTO, and the remaining 13 variables are variables constructed by Jaffe and Trajtenberg. The second file Cite75_99 contains a pairwise list of all utility patents granted between 1975 and 1999 and the patents they cited back to 1963; the file contains over 16 million records.

Joining the Pat63_99 and Cite75_99 file, 3 distance measures (for each patent) were calculated. For now, a simple point system assigns a distance depending on how far the citing patent, p_n, is from the cited patent. If the citing patent shares the same class, c_1, as the cited patent then the distance is small (distance = 1); if the patents share the same subcategory, c_2, then the distance is larger (distance = 2); if the patents share the same category, c_3, then the distance is still larger (distance = 3); if the patents come from different categories, c_4, then the distance is still larger (distance = 4).

Again, a patent distance score measures the breadth of expertise required to implement the patent. Long distance connections from different patent categories are given the highest points; the distance between $d(p_n, p_m)$ in figure 3 is 4, since it connects knowledge from a completely different category. This long distance connection is equivalent to a patent in the category Chemical citing a patent in Computers.

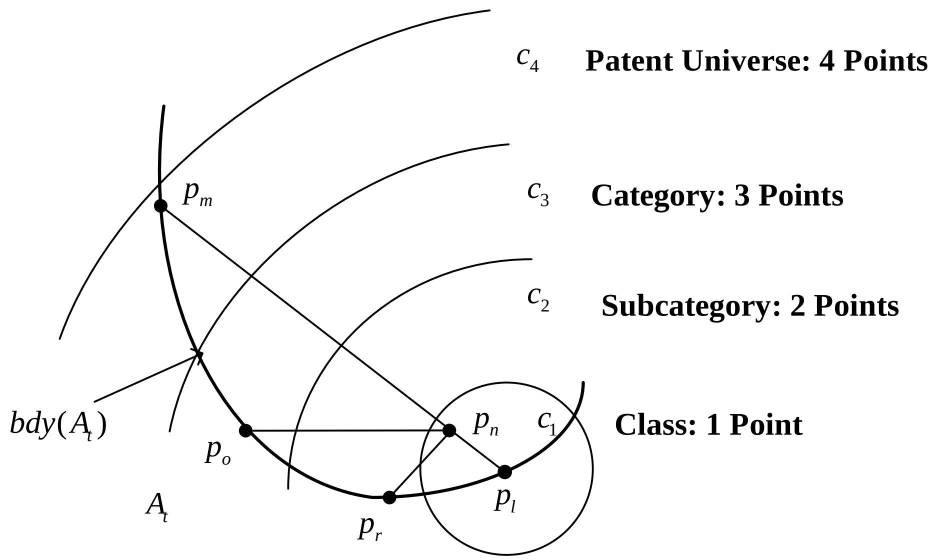


Figure 3. Distance Point System.

The first measure of distance is the maximum distance which scores each patent based on the farthest distance among all the citations it made. The max distance adjusts for patents that are making citations that are useless or unnecessary. The max distance for the citing patent, p_n, in figure 3 is equal to 4.

$$\text{Max Distance}(p_n) = Max(d_i(p_n, p_i))$$
$$for \ i = 1, \ldots, N$$
N : the number of citations made by citing patent p_n
i: each different cited patent

The second measure is the total distance for each patent and is measured by calculating the distance between each citing and cited patent and then aggregating. For example, the total distance for the patent in figure 3 is equal to 10. This measure rewards patents by giving at least 1 point for each citation and more points for citations in farther patent classes.

$$\text{Total Distance}(p_n) = \sum_{i=1}^{N} d_i(p_n, p_i)$$
N : the number of citations made by citing patent p_n
i: each different cited patent

The third measure is the average distance for each patent and is measured by calculating the distance between each citing and cited patent, aggregating, and dividing by the number of citations. For example, the average distance for the patent in figure 3 is equal to 2.5. This measure rewards patents that make long distances per citation;

thus if all the citations in figure 3 were made within the same category (and outside the subcategory) the average distance for this patent would be 3.

$$\text{Average Distance}(p_n) = \left(\sum_{i=1}^{N} d_i(p_n, p_i) \right) / N$$

N : the number of citations made by citing patent p_n

i : each different cited patent

PRELIMINARY EMPIRICAL RESULTS

Technological Distances and Team Size

Jones (2005) has already provided evidence that team size is increasing over time; the goal now is explain why this is happening. Since this paper proposes that technological distance measures the breadth of expertise needed to implement an idea at least 2 properties of technological distance should hold. First, it should be true that team size is positively related to technological distance. The reason for this is that ideas that span large technological distances span a large area of expertise and thus require larger teams to implement. Table 1 and table 2 show that this property holds across all 3 technological distance measures in 1975 and 1995 respectively. Additionally, figure 3 and figure 4 show the graphical results from table 1 and 2 for the average total distance (ATD) measures for 1975 and 1995; and together they show a shift upwards in average total distance graph across all team sizes.

Table 1: Team Size for U.S. R&D Workers and Average Distances 1975

Team Size	Total Distance	Max Distance	Average Distance	Observations
1 (solo)	7.00	2.50	1.86	35008
2	7.57	2.58	1.92	15258
3	7.90	2.62	1.96	6112
4	7.65	2.57	1.98	2224
5	7.79	2.64	2.01	827
6	7.62	2.56	1.91	349
7	7.38	2.65	2.11	141
8	7.30	2.40	1.84	84
9	5.96	2.34	1.83	50
10	6.71	2.54	2.04	10

[a] Excluded from the table are team sizes greater than 10

Table 2: Team Size for U.S. R&D Workers and Average Distances 1995

Team Size	Total Distance	Max Distance	Average Distance	Observations
1 (solo)	19.34	2.99	1.94	53761
2	22.79	3.07	2.02	23688
3	23.28	3.10	2.05	21007
4	23.37	3.12	2.07	11460
5	23.57	3.11	2.08	5865
6	24.66	3.18	2.12	3082
7	25.80	3.14	2.11	1568
8	22.97	3.02	2.07	890
9	30.02	3.21	2.15	481
10	27.52	3.29	2.05	297

[a] Excluded from the table are team sizes greater than 10

Secondly, it should be true that on average solo R&D workers have smaller distances than teams of R&D workers. The reason for this is because solo R&D workers can only implement ideas completely within their own area of specialization. The evidence indicates that solo R&D worker distances are on average smaller than the distances for teams of R&D workers.

Figure 3. Average Total Distance (ATD) and Team Size 1975

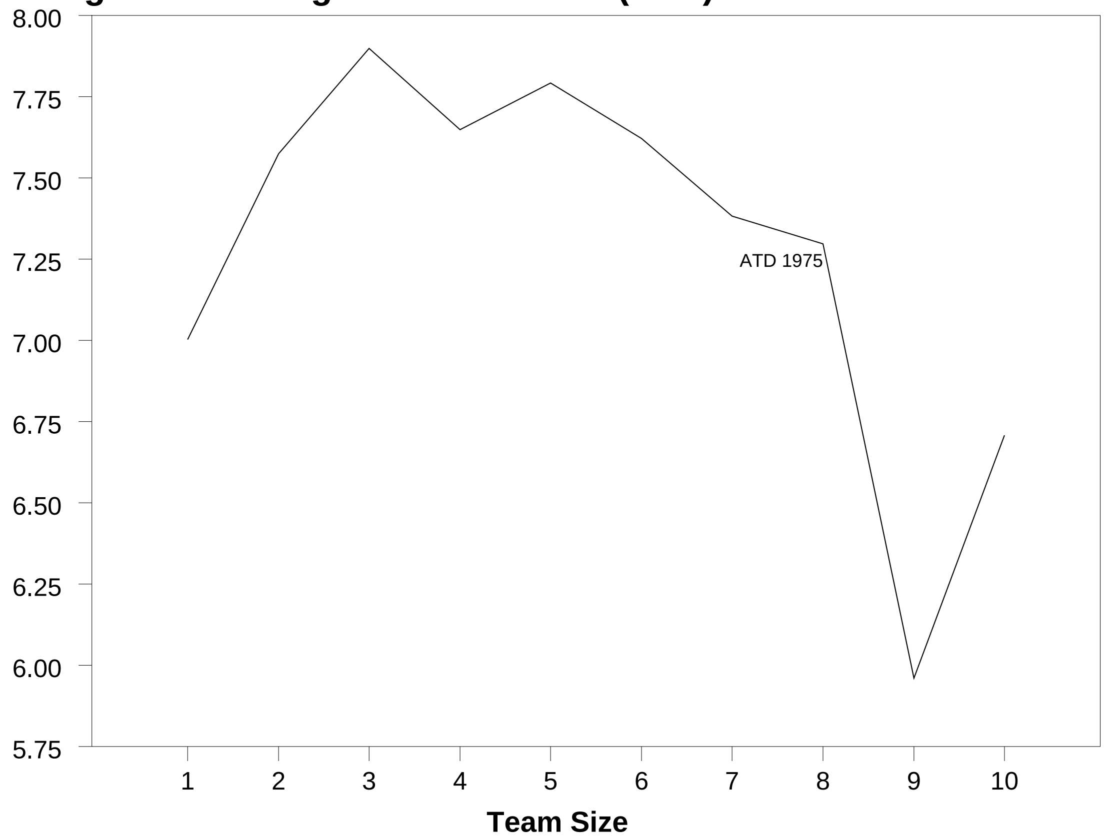
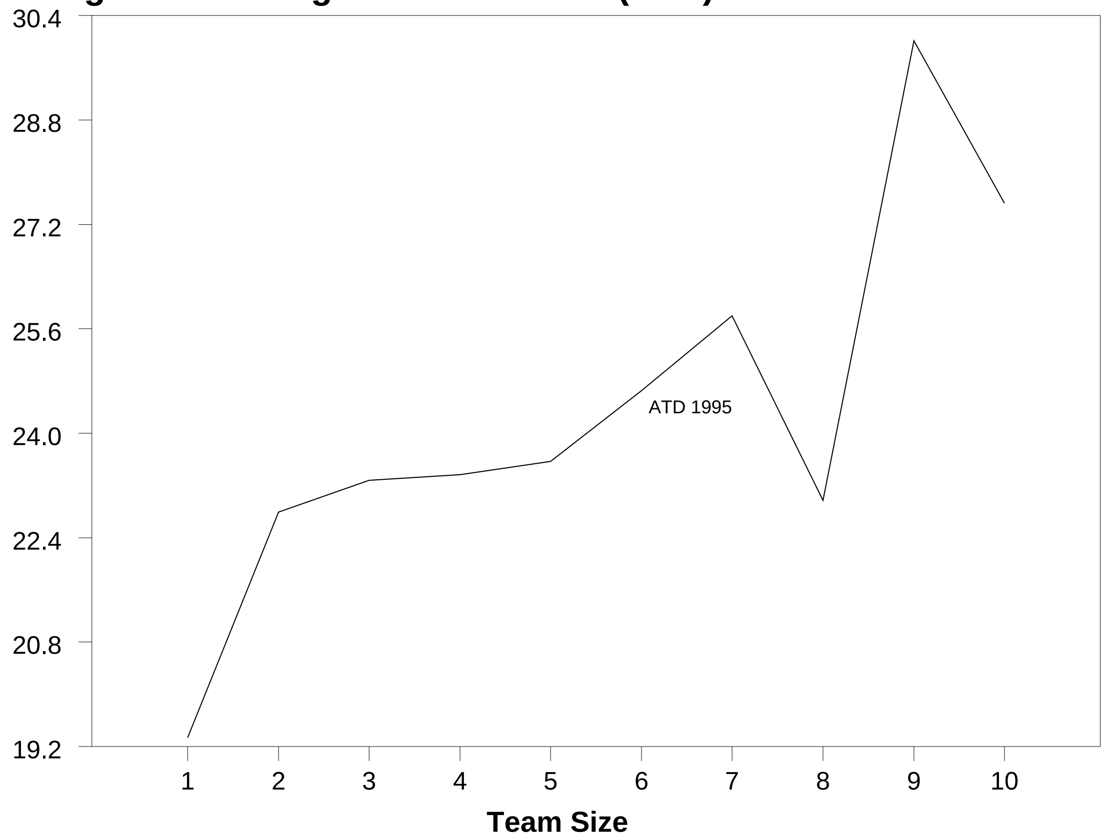


Figure 4. Average Total Distance (ATD) and Team Size 1995



Technological Distances over Time

Technological distances were constructed to determine if R&D workers were forming larger teams because they were struggling to implement larger ideas that were outside their own area of expertise. The time series data for technological distances are constructed by taking all the patents granted to U.S. R&D workers for a particular year and calculating the average distance for that year of data. Figures 5 shows average annual max distances across the top 4 G-7 countries. The average maximum distance for U.S. R&D workers in 1975 of 2.5 indicates that many patents connected knowledge that was within the same category but outside the subcategory. By 1989 the average maximum distance reached 3.0 meaning that many patents connected knowledge that was outside their category. This indicates that patents are making more citations to knowledge that is farther away evidence that R&D workers are getting bigger ideas over time.

The graphical evidence shows that technological distances are increasing across G-7 countries too. The United States is the leader across all G-7 countries with the UK, Canada, and France following; Japan, Italy, and Germany (table 3) are lagging. Again this evidence suggests that ideas are getting bigger over time and across countries. The trend is apparent across industrialized countries, although only the average max distance is shown the trend is true for both the average annual total distance and the average annual average distance across G-7 countries (suppressed for space considerations). Furthermore, a similar investigation was done across the 4 Asian Tigers: Singapore, Hong Kong, South Korea, and Taiwan and the results for these countries resemble the trend for Japan; although with substantial volatility. Among the 4 Asian Tigers Singapore has the highest technological distance, followed by Hong Kong, South Korea, and Taiwan.

Figure 5: Average Max Distance for Top 4 G7 Countries

Annual Data by Application Year 1975 - 1995

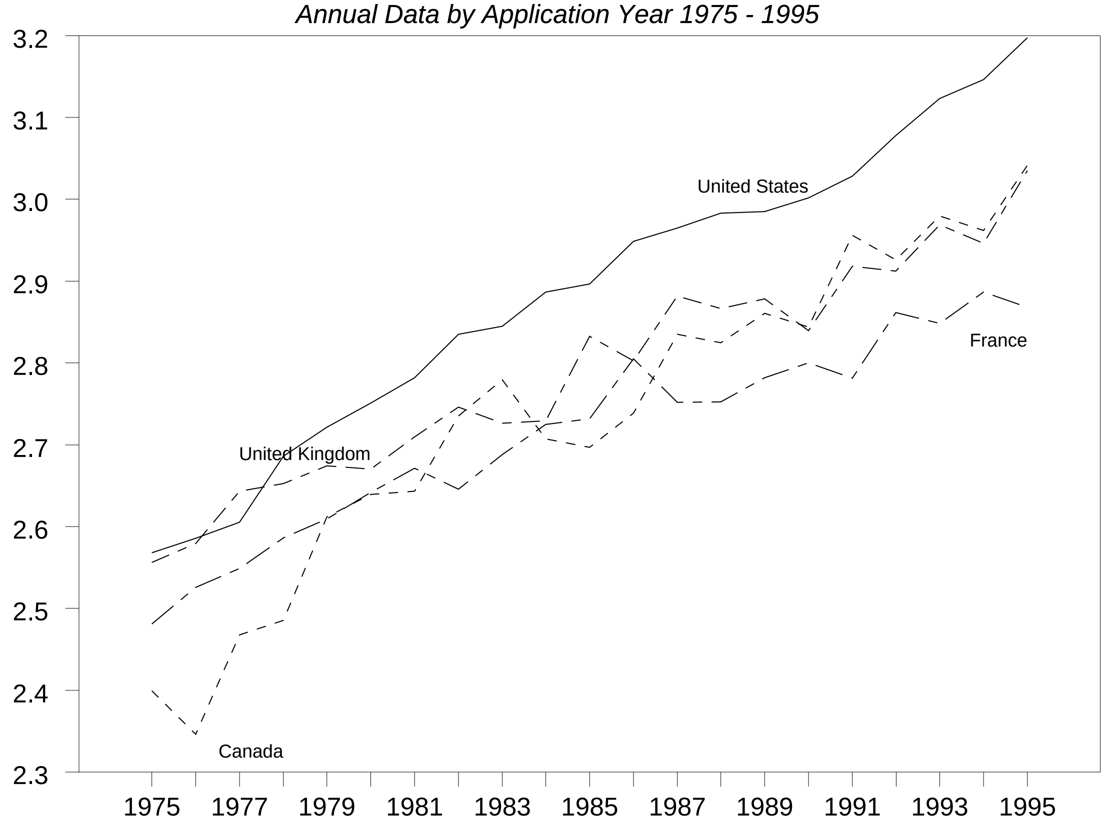


Table 3: Descriptive Statistics for Average Maximum Distance (AMD)

Country	Mean	Std Error	Minimum	Maximum
US	2.88	0.21	2.47	3.21
UK	2.78	0.16	2.40	3.06
Canada	2.73	0.23	2.24	3.09
France	2.70	0.15	2.30	2.95
Japan	2.66	0.12	2.40	2.85
Germany	2.65	0.14	2.31	2.91
Italy	2.58	0.18	2.07	2.85

[a] Sorted descending by Mean.

[b] Quarterly data for G7 Countries 1975 – 1999 by Grant Year

The data presented here indicates that technological distances are increasing in the United States and that this increase partially explains the increase in team size documented by Jones over the same period; thus, providing a joint reason for increasing team size. In addition, the data reveals a specialization paradox; a priori, one would think that solo R&D worker distances should decline over the sample in keeping with the hypothesis of increased specialization; however, Tables 1 and 2 indicate that solo R&D workers are keeping up (relatively) with team R&D worker distance. This means that solo R&D workers and are conceiving and implementing larger ideas over time.

CONCLUSIONS

Increasing knowledge puts an increased learning burden on successive generations of R&D workers. One way to deal with this burden is for R&D workers to increasingly specialize and form teams to implement their new ideas. However, if R&D workers are getting bigger ideas over time then larger teams are need too. This paper uses the Olsson model of ideas to measure the sizes of idea for United States patents granted from 1975 – 1995. Preliminary results show that idea size is increasing over time signaling that increased specialization only partially explains increasing team size.

REFERENCES

Jaffe, A. B. and Trajtenberg, M. (2002). Patents, Citations, and Innovations. Cambridge, MA: MIT Press.

Jones, B. F. (2005). "The Burden of Knowledge and the 'Death of the Renaissance Man': Is Innovation Getting Harder?," NBER Working Paper no. 11360.

McCloskey, D. (2000). How to Be Human* Though an Economist. Ann Arbor: The University of Michigan Press.

Olsson, Ola (2005). "Technological Opportunity and Growth," Journal of Economic Growth 35-57.

Olsson, Ola (2000). "Knowledge as a Set in Idea Space: An Epistemological View of Growth," Journal of Economic Growth 253-275.