Inventing by Combining Pre-Existing Technologies:

Patent Evidence on Learning and Fishing Out

Matthew S. Clancy*

Abstract

I develop a model of innovation where new technologies are combinations of pre-existing technological components. The model captures two opposing forces. The best ideas are used up (knowledge is exhaustible). However, as firms learn which technologies can be combined, new ideas become feasible (knowledge accumulates). I test the model with more than 80 years of US patent data. Technological components are proxied by 13,517 patent office technology classifications. These are reused and recycled in 10,000 distinct three-component sets. Consistent with a learning/fishing-out dynamic, I show patenting in one set of components is correlated with a subsequent increase in similar patents (sharing two of three components), but a subsequent decrease in identical patents (sharing all three components). I use patent renewal data to show my results are not driven by changes in demand for various technology bundles. My results suggest the positive impact of learning on subsequent patenting is larger than the negative impact of fishing out.

Keywords: Innovation, Patents, Combinatorial growth, Spillovers, R&D

JEL Codes: O31, O34

* Matt Clancy (matthew.clancy@ers.usda.gov) is an economist at the USDA Economic Research Service, Washington DC.

Opposite forecasts for the outlook of innovation currently coexist. In one view, rapid innovation lies ahead: artificial intelligence will reshape the economy (McAfee and Brynjolfsson 2014, Bostrom 2014) as we take to other planets (Vance 2015) and use genetic engineering to control our evolution (Doudna and Sternberg 2017). But in another view, continuous innovation is an exception, and stagnation is the rule. We have already discovered all the good ideas and, as a consequence, innovation is likely to slow and stall (Cowen 2011, Gordon 2016). These views differ in their evaluations of two opposing dynamics in innovation. The first emphasizes innovation as a primarily cumulative process: as we learn more, the applications worth exploring multiply. In this paper, I refer to this as the learning effect. The second view emphasizes that knowledge is more like a finite natural resource extracted by research. This is frequently referred to as the fishing out effect. The outlook for innovation depends on which of these features dominates. Are we fishing out the stock of ideas faster than learning "restocks" it? This is an empirical question and this paper develops a novel methodology to answer it.

Psychologist of creativity Keith Sawyer writes creativity is "a new mental combination that is expressed in the world" (Sawyer, 2012, pg. 7). My starting point is a model of innovation wherein ideas are new combinations of pre-existing technological components. Consider the internal combustion engine as a representative example. While it is a single idea, it can also be viewed as a combination of constituent components: pistons, crankshafts, flywheels, and so on. Each of these components existed prior to the engine, and the engine's discovery required assembling pre-existing constituent components into a combination not previously known (Dartnell 2014, pg 201).

This way of thinking about discovery has a long history. Mathematician Henri Poincaré argued, "[T]o create consists precisely in not making useless combinations and in making those which are useful and which are only a small minority." (Poincaré 1913, pg. 386). Abbott Payson Usher's A History of Mechanical Inventions noted, "Invention finds its distinctive feature in the constructive assimilation of preexisting elements into new syntheses, new patterns, or new configurations of behavior" (Usher, 1929, pg. 11). Schumpeter described the essence of enterprises and entrepreneurship to be "the carrying out of new combinations" (recounted in Weitzman 1998, pg. 335). This perspective has also been articulated in formal models by Weitzman (1998), Olsson and Frey (2002), Simonton (2004), Olsson (2005), Feinstein (2011), Ghiglino (2012), and Akcigit, Kerr and Nicholas (2013).

A straightforward interpretation of "fishing out" follows from combinatorial models of innovation. This paper assumes a given combination has a fixed number of distinct applications (i.e., there are only so many ways to combine pistons, crankshafts, flywheels, and so on to obtain something novel and useful), so that the stock of ideas is finite and R&D draws it down.[1] Combinatorial models can also model the cumulative nature of knowledge. My model is most closely related to the concept of "clumps" in Arthur 2009, in which some components (such as pistons and crankshafts) are understood to go together naturally because they "repeatedly form subparts of useful combinations" (Arthur 2009, pg. 70).[2]

To briefly illustrate the thrust of this paper, consider three technological components, *x*, *y*, and *z*, that may be combined into a new idea *xyz* with some probability and at some cost. The

[1] Weitzman (1998) and Akcigit, Kerr, and Nicholas each incorporate a variant of fishing out effect.

[2] Others have modeled the knowledge associated with combinatorial innovation as arising from the discovery of new components (Weitzman 1998, Akcigit Kerr and Nicholas), or the discovery of new combinations that can be repeated to diminishing effect (Akcigit, Kerr and Nicholas 2013), or the discovery of combinations or ideas that bridge distant technological spaces (Olsson 2005 and Feinstein 2011) and reveal the quality of "nearby" ideas (Jovanovic and Robb 1990, Kauffman, Lobo and Macready 2000, Auerswald et al. 2000).

combination is more likely to succeed if researchers can observe prior instances where the components have been combined usefully. This knowledge is modeled by how many times each of the *pairs* of components (*xy*, *xz*, and *yz*) have been combined successfully. However, the exact combination of components *xyz* can only be "discovered" a finite number of times.

In this model, every new idea affects future innovation through both learning and fishing out effects. Suppose a fourth technological component, *w*, is also available. If *xyz* is a successful combination, researchers observe an instance of the pairs *xy*, *xz*, and *yz* being combined. This increases the probability that combinations such as *wxy*, *wxz*, and *wyz* will also succeed, as these combinations make use of the same pairs. This is the positive learning effect, where every discovery makes similar research more attractive. At the same time, one instance of the precise combination *xyz* has been used up by its discovery. This is the negative fishing out effect.

There is a long line of empirical papers in this literature. Ideas are usually proxied by academic papers or patents, and the citations they make to antecedents in different fields determine the extent of recombination in an idea. A few papers (e.g., Fleming 2001 and Akcigit, Kerr, and Nicholas 2013) instead use the technological classifications directly assigned to patents as proxies for technological components. This is the approach I take.

Much of this literature has looked for correlations between the combinatorial properties of patents/papers and their subsequent citations. Because citations can be interpreted as proxies for knowledge flows, this line of literature can also be interpreted as providing some evidence on the learning effect. Patents/papers that receive more forward citations are ideas from which many future researchers learned something important. A long stream of studies[3] has generally found

[3] See Hall et al. (2001), Fleming (2001), Schoenmakers (2010), Schilling (2011), Nemet (2012), Akcigit, Kerr, and Nicholas (2013), Kaplan and Vakili (2015). Nemet and Johnson (2012) is an

that recombination is associated with more citations, and therefore (perhaps) learning. Conversely, the extent to which familiar combinations do *not* generate new citations could be read as evidence that these technological domains are fished out. Fleming (2001) provides more direct evidence on fishing out by showing a patent is less likely to be highly cited if its exact combination of subclasses has been patented more often.

This paper differs from the above in several respects. My unit of observation is a specific combination in a particular year, not a paper or patent. My dependent variable is not citations, but the number of patent applications in a given year with a particular combination of technological components, including years in which no patent applications for a given combination are filed (the majority of cases). By looking at the factors correlated with the number of applications with a particular combination, I can measure the empirical import of various variables associated with the combination.

Moreover, my proxies for learning and fishing out allow me to identify these effects separately. I assume a combination is fished out by identical combinations. For example, the combination *xyz* is only fished out by patents assigned the exact set *xyz* (this is also how Fleming 2001 measures fishing out). However, the learning effect is driven by the number of patents using various pairs of elements in a set (i.e., patents containing any of *xy*, *xz*, and *yz*). This gives me differential variation in learning and fishing out, which I use to estimate their relative magnitudes. The chief contribution of the paper is demonstrating that the learning effect exceeds the fishing out effect.

example of a contrary finding. Uzzi et al. (2013) and Keijl et al. (2016) suggest it is not necessarily the total amount of recombination that matters, but that an atypical combination was made within a familiar context.

However, this exercise is only useful to the extent that the underlying model and causal interpretations are correct. My second contribution is providing novel evidence to support the model. I derive and find empirical support for five hypotheses suggested by this paper's model of combinatorial innovation. Additionally, I use patent renewal data to rule out an alternative interpretation of the data, that my measure of "learning" is merely proxying for lagged changes in demand for different technologies.

Third, my use of technology subclasses improves on earlier work by aggregating up to the mainline class. This ensures that technology classifications are non-nested, exhaustive, and comparable. Aharonson and Schiling 2016 have recently explored a similar approach as applied to maps of the technological landscape.

The layout of this paper is as follows. In section 1, I set up my model and supplies four of the five hypotheses that will be tested. In section 2, I describe the historic US patent data I use in the paper's empirical application. Section 3 describes my econometric methodology. Section 4 presents my results, and evaluates how well they support the four hypotheses developed in section 1. Section 5 extends the analysis by introducing patent renewal data to both test a fifth hypothesis and to eliminate the alternative hypothesis that my results are driven by demand-side factors. Section 6 compares the size of the learning and fishing out effects. Section 7 concludes with a summary of the paper's contributions and some suggestions for future research.

1. A Model of Combinatorial Innovation

1.1 Model

This section presents a three-step model of R&D and patenting. The first part is a combinatorial model of the R&D process. The second is a simple model of how firms decide which R&D projects to initiate. The third part combines the first two to derive predictions about which ideas are patented.

We begin with a model of the R&D process. There exists a set Q of pre-existing technological components. Ideas are subsets of Q with at least two components. Let a subset be denoted by i. The purpose of R&D is to determine if a set of components results in a viable invention, where "viable" means simply that the invention works, in the sense of meeting the desired technical specifications.

The viability of an invention is a function of how its constituent components interact with each other. I define a scalar measure called *affinity* that measures the state of knowledge about how two components can be usefully combined. Let the affinity at time t between a pair j of components be denoted A_{jt}. There is an unobserved "true" affinity that is time-invariant and measures the true utility of combining technological components; A_{jt} converges to this "true" value as information about how components may or may not be combined accumulates. In particular, A_{jt} increases with the number of examples of viable ideas using component-pair j and decreases with the number of examples of unviable ideas using component-pair j.

Inventions are more likely to be viable (from the perspective of researchers) if their components have high affinity for each other. Let $\overrightarrow{A_{it}}$ denote the vector of affinities of all pairs of set i's components at time t (there will be $n(n\text{-}1)/2$ pairs of components if the set i has n components). The probability that an idea with set i is viable is a function of the affinities between its components:

$$\Phi\left(\overrightarrow{A_{it}}\right) \equiv \Pr\left(i \text{ viable}\right) \tag{1}$$

where I assume:

$$\frac{\partial \Phi}{\partial A_{jt}} \geq 0 \forall A_{jt} \in \overrightarrow{A_{it}} \tag{2}$$

That is, the probability an idea is viable is (weakly) increasing in the affinity of the components that comprise the idea. As to the cross derivative of some A_{jt} and A_{-jt}, I consider two extreme cases:

Perfect substitutes: $$\Phi\left(\overrightarrow{A_{it}}\right) = \Phi_S\left(\sum_{A_{jt} \in \overrightarrow{A_{it}}} A_{jt}\right)$$

Perfect complements: $$\Phi\left(\overrightarrow{A_{it}}\right) = \Phi_C\left(\min\left(\overrightarrow{A_{it}}\right)\right)$$

An example will help clarify this model of the R&D process. Suppose we are inventing an internal combustion engine by combining flywheels, pistons, crankshafts, and myriad other pre-existing technological components from Q. We will consider the invention a viable one if our array of components lets us pour fuel in and get rotational motion out.

Some of the components have a high affinity for each other. Pistons and crankshafts have a long history of joint use (for example, in waterwheels – Dartnell 2014) so that researchers know how to usefully couple them. Potter's wheels have long used flywheels to convert the uneven rotational energy of a crankshaft into smooth motion, so that these components also have high affinity for each other. Note that in this model, the knowledge that was drawn on to build the internal combustion engine came from domains that used only a small subset of the internal combustion engine's components. This is reflected in the model by affinity, a measure of knowledge about how just two components interact, where knowledge can be drawn from any domain using those components.

Other components may not have high affinity. For example, suppose researchers are unsure if there will be a problem using a flywheel with a piston. These two components have a low affinity for each other. How much does this weaken the expected viability of the internal combustion engine? If every component interacts with every other component, then this missing knowledge is very important. An invention is only as viable as its weakest link. If only some components need to interact, then the engine can be designed so that pistons and flywheels do not directly interact. In this case, it is not very important that researchers don't know how to usefully combine them. Perfect complements defines the first case, perfect substitutes the second.

We now turn to the second part of the model, the R&D decision. New ideas are created by myopic profit-seeking firms and firms may conduct R&D on one idea per period. Research on an idea *d* with set *i* in period *t* costs K_{dt}, and at the end of the period, firms learn if the idea is viable. Viable ideas are patented, and the inventing firm obtains a flow of rents with expected present discounted value V_{dt}. There is a new realization of V_{dt} and K_{dt} in each period:

$$V_{dt} = v_i + u_t^k + e_{dt}^k \tag{3}$$

$$K_{dt} = k_i + u_t^k + e_{dt}^k \tag{4}$$

where e_{dt}^h is an idiosyncratic error for idea *d* using set *i*, for $h = v, k$. Note that v_i and k_i are set-specific, so that some ideas may have persistently higher (or lower) draws across all years, and u_t^h is time-specific, so that some years may have persistently higher (or lower) draws across all ideas. However, I assume draws of V_{dt} and K_{dt} are otherwise uncorrelated. This is a rather strong assumption, but a necessary one for my model's identification. In the extensions of my model in section 6, I present evidence on the validity of this assumption and discuss how relaxing it affects the interpretation of my empirical results.

Firms know V_{dt} and K_{dt} when deciding whether to conduct R&D, but not whether an idea is viable (which is determined by the previously discussed R&D process). Once an idea has been attempted, no other firm attempts the idea: it is either patented or discovered to be unviable, and this is public information. The value V_{dt} is related to, but distinct from, the viability of an invention. Ideas that are not viable have zero value: they are fatally flawed and not useful to any buyer. Nonetheless, viable ideas may not be particularly valuable. Many ideas that are technical breakthroughs do not end up being highly valued by society, at least initially (e.g., the Segway, Google Glass).

Every period, new values of K_{dt} and V_{dt} are drawn, representing changes associated with the cost of R&D and the value of individual inventions. There is free entry into the innovation game, and in each period a single[4] firm attempts each idea satisfying:

$$V_{dt}\Phi\left(\overrightarrow{A_{it}}\right)-K_{dt}\geq 0 \tag{5}$$

Equation (5) asserts that R&D is attempted whenever it has (weakly) positive expected value. This condition can be rearranged to yield:

$$K_{dt}/V_{dt}\leq \Phi\left(\overrightarrow{A_{it}}\right) \tag{6}$$

We are now in a position to turn to the third part of the model, a discussion of which ideas are patented. Define $\Lambda_{it}\left(K/V\right)$ to be the cumulative density function of K/V (recall both of these variables are themselves random). I assume a specific combination i can be tried $N_i \geq 1$ times, where N_i gives the number of combinations sufficiently distinct to be patentable. Let M_{it} denote the number of ideas that have been attempted using set i up through period t. Finally, let

[4] We may assume firms enter each period in a random ordered sequence and may claim any unclaimed idea upon entry. Claims are public knowledge and firms complete their R&D in the same sequence, and so no two firms will ever attempt the same idea.

y_{it} denote the number of patent applications of ideas with set *i* in period *t*. The expected number of patents in a given period is:

$$E[y_{it}] = (N_i - M_{it}) \times \Lambda_{it}\left(\Phi\left(\overrightarrow{A_{it}}\right)\right) \times \Phi\left(\overrightarrow{A_{it}}\right) \quad (7)$$

The first term in equation (7) is the number of untried ideas remaining in period *t*, the second term is the probability each such idea will be attempted (derived from equation (6)), and the third term is the probability each attempted idea will be viable (and therefore patented).

Equation (7) succinctly captures a range of possible drivers of innovation: pull factors, push factors, and shocks to the state of knowledge. The example of the internal combustion engine's grandfather, the steam engine, provides a useful illustration. The engine was technically feasible once the underlying components became available, as attested by Hero of Alexandria's working design in the 1st century A.D. (Mokyr, 1990, pg. 22). However, the engine was not widely used until the design was rediscovered and refined in the 1700s in Great Britain by Thomas Newcomen and later by James Watt. Economic historians seeking to understand the industrial revolution have proposed a large range of potential explanations (far more than will be used here for illustration) for why the engine was developed at this time and place.

For example, Allen (2009) argues the price of coal was relatively low compared to labor in Britain, which made the steam engine profitable to use in Britain but not elsewhere. In the context of the model, this can be represented as a high draw of V_{dt}. In contrast, Meisenzahl and Mokyr (2011) point to Britain's large supply of skilled artisans and craftsman, who assisted inventors or undertook invention themselves. This human capital advantage made R&D easier and less costly than elsewhere, and we can model this as a low draw of K_{dt}. Holding fixed

$\Phi\left(\overrightarrow{A_{it}}\right)$, the engine might have been invented in any period with a sufficiently small draw of *K/V*.

An alternative explanation sees shocks to the state of knowledge as being fundamental. Wootton (2015) presents evidence that it was Newcomen's encounter with Denis Papin's description of the mechanics of pressure cookers that provided him the crucial insights he needed to improve the steam engine (Wootton 2015, pgs. 499-508). In the context of the model, developments in a distinct but related invention (the pressure cooker), provided important new information about how to usefully combine pre-existing components in the steam engine. An increase in A_{jt} led to an increase in the probability the steam engine was viable ($\Phi\left(\overrightarrow{A_{it}}\right)$) such that draws of *K/V* that were previously insufficient to support R&D became sufficient.

1.2. Empirical Application

I summarize equation (7) in reduced form as:

$$E\left[y_{it}\right]=\tilde{\Psi}\left(i,t,\overrightarrow{A_{it}},M_{it}\right) \tag{8}$$

That is, the expected number of patents using mainline set i is a function of four variables: the set *i* (encompassing set specific values of N_i, v_i, and k_i), time *t*, the affinity $\overrightarrow{A_{it}}$ between pairs of components in the patent, and the number of attempted ideas M_{it}. In particular:

$$\frac{\partial\tilde{\Psi}}{\partial A_{jt}}=\left(N_{it}-M_{it}\right)\left\{\Lambda_{it}'\Phi\left(\overrightarrow{A_{it}}\right)+\Lambda_{it}\right\}\partial\Phi\left(\overrightarrow{A_{it}}\right)/\partial A_{jt}\geq 0 \tag{9}$$

$$\frac{\partial\tilde{\Psi}}{\partial M_{it}}=-\Lambda_{it}\Phi\left(\overrightarrow{A_{it}}\right)\leq 0 \tag{10}$$

Note that because Λ_{it} and $\Phi\left(\overrightarrow{A_{it}}\right)$ correspond to probabilities, $\tilde{\Psi}$ is bounded from above by N_i and from below by 0. For sufficiently high levels of A_{jt} and M_{it}, $\tilde{\Psi}$ approaches its bounds and so the partial derivatives given in equations (9) and (10) must decline towards 0 as well.

I do not actually observe the components of ideas. Instead, I observe their proxies in the form of the patent office's technology classifications (discussed in the next section), which are called **mainlines**. Neither do I observe M_{it} or A_{jt}. Instead, I observe two related proxies. Let m_{it} denote the number of times mainline-set *i* has been assigned to a patent granted prior to period *t*, and let a_{jt} denote the number of times mainline-pair *j* has been assigned to a patent granted prior to period *t*. For example, if three patents have been granted with mainline sets *xyz*, *xyz*, and *wxy*, then for mainline set *xyz*, $m = 2$, and for the mainline set *wxy*, $m = 1$. Meanwhile, for pair *xy*, $a = 3$, for pairs *xz* and *yz*, $a = 2$, for pairs *wx* and *wy*, $a = 1$, and for *wz*, $a = 0$.

Besides the measurement error imposed by using mainlines as proxies, these variables are imperfectly correlated with the "true" M_{it} or A_{jt} because I only observe viable ideas. Mainline combinations that are unviable do not result in patents, but do reduce the number of ideas to be tried and may impact A_{jt} if researchers observe unviable ideas. For example, if *wxz* is inviable, this may reduce researchers beliefs that *w* and *x* can be usefully combined. Moreover, information about the affinity of a pair may come from other sources, such as successful but unpatented innovations, or related scientific work.

Accordingly, I supplement m_{it} and a_{jt} with the *age* of a mainline combination. Failed attempts to combine components and other information accumulates over time, starting from the moment the mainlines first become available to researchers. Thus, *age* measures the time elapsed since the mainline-set first became available. In my empirical model, I end up estimating the following models:

Perfect Substitutes:

$$E\left[y_{it}\right]=\Psi_S\left(\beta_1\sum_{j\in i}a_{jt}+\beta_2\sum_{j\in i}a_{jt}^2+\phi_1 m_{it}+\phi_2 m_{it}^2+\gamma_1 age+\gamma_2 age^2+X'\theta\right) \tag{11}$$

Perfect Complements:

$$E\left[y_{it}\right]=\Psi_C\left(\beta_1\min_{j\in i}\left(a_{jt}\right)+\beta_2\min_{j\in i}\left(a_{jt}^2\right)+\phi_1 m_{it}+\phi_2 m_{it}^2+\gamma_1 age+\gamma_2 age^2+X'\theta\right) \tag{12}$$

where *X'θ* is a set of controls potentially including time trends and mainline-set fixed effects (as in equation (8)), and Ψ_S and Ψ_C are non-linear functions with $\Psi'_S>0$ and $\Psi'_C>0$ (discussed in section 3). Both the perfect substitutes and perfect complements framework can be thought of as approximating the relationship between A_{jt} and a_{jt} by the quadratic $A_{jt}=\beta_1 a_{jt}+\beta_2 a_{jt}^2$. In the perfect substitutes framework, the sum of A_{jt} for each pair in a set is $\sum_{j\in i}A_{jt}=\beta_1\sum_{j\in i}a_{jt}+\beta_2\sum_{j\in i}a_{jt}^2$, while in the perfect complements framework, the minimum A_{jt} of all the pairs in an idea is $\min_{j\in i}\left(A_{jt}\right)=\beta_1\min_{j\in i}\left(a_{jt}\right)+\beta_2\min_{j\in i}\left(a_{jt}^2\right)$. I use the quadratic specification for a_{jt} and m_{jt} not because I believe the true relationship is quadratic, but because the quadratic structural form allows me to make falsifiable predictions. In particular, I use equations (11) and (12) to test four hypotheses:

Hypothesis 1 (marginal impact of learning): $\beta_1+2\beta_2 x\geq 0$, where $x=\sum_{j\in i}a_{jt}$ or $x=\min_{j\in i}(a_{jt})$, over the domain of the observations.

This follows directly from equation (9).

Hypothesis 2 (learning upper bound): $\beta_1>0$ and $\beta_2<0$.

This implies there is an upper bound on the returns to knowledge, and that the marginal returns to additional instances of successful combination of a pair fall to zero at some point. These hypotheses jointly imply another test:

Auxiliary Hypothesis (learning): $-\beta_1 / 2\beta_2$ is greater than the maximum observation of $\sum_{j \in i} a_{jt}$ or $\min_{j \in i}\left(a_{jt}\right)$.

Note that $-\beta_1 / 2\beta_2$ is the turning point of the quadratic. If hypothesis 2 is correct, then it is the maximum value, and increasing affinity beyond this point has a negative marginal impact. If my model is correct, there should be no negative marginal impact, and Hypothesis 1 requires that this turning point be outside the range of observations. These hypotheses jointly test that the relationship between affinity and the number of patents is well approximated by the left-hand side of an inverted-U shape. To these hypotheses I add:

Hypothesis 3 (marginal impact of fishing out): $\phi_1 + 2\phi_2 m_{it} \leq 0$ over the domain of the observations.

This follows directly from equation (10).

My model also implies a lower bound for the number of patents, and that the impact of fishing out falls to zero once $M_{it} = N_i$. However, this does not imply a testable hypothesis, because my data necessarily has a lower bound of 0 (I do not observe negative patent counts) and the functional form of Ψ_S and Ψ_C also has a lower bound of 0. Therefore, so long as hypothesis 3 is correct, increases in m_{it} will necessarily have diminishing marginal impact above a certain level.

Hypothesis 4 (marginal impact of age): Either $\gamma_2 < 0$ or $\gamma_1 + 2\gamma_2 age \leq 0$ over the domain of observations.

Hypothesis 4 posits that the impact of age is either negative, or increasing but bounded over some range. In my model, the variable *age* proxies for unobserved information that accumulates over time. This unobserved information could include research projects that do not result in viable ideas, and which are unpatented. If these attempts fish out potential combinations, then *age* behaves like m_{it} and the marginal impact of *age* should be negative ($\gamma_1 + 2\gamma_2 \leq 0$) as under hypothesis 3. Conversely, if these attempts yield useful information about the affinity of pairs, or in general if information outside the patent system provides information about the affinity of pairs, then age primarily operates through its effect on affinity. If *age* primarily proxies for information that serves to increase the affinity of a pair, then its marginal impact may be positive, but bounded ($\gamma_2 < 0$). What hypothesis 4 rules out is a constant or increasing marginal impact of *age* over the domain of observations.

2. Data

2.1 The Patent Classification System

My data draws on the full set of US utility patents granted between 1836 and 2009: 7.6 million patents. The US Patent and Trademark Office (USPTO) has developed the US Patent Classification System (USPCS) to organize patent and other technical documents by common subject matter. Subject matter can be divided into a major component, called a class, and a minor component, called a subclass. USPTO (2012) states, "A class generally delineates one technology from another. Subclasses delineate processes, structural features, and functional features of the subject matter encompassed within the scope of a class." Subclasses are a natural candidate for the building blocks of combination, out of which researchers build new ideas.

Using patent subclasses as proxies for the building blocks of ideas has many advantages over plausible alternatives, such as the words used in a patent document or citations to prior art. Unlike text or citations, patent classifications are chosen by an ostensibly disinterested party, namely the patent examiner. Classifications have no special legal standing and are not generally of interest to patent applicants (and therefore not chosen strategically). Instead, they are chosen to facilitate searches by future parties who wish to verify that new applications are, in fact, novel. Furthermore, the classification system is updated over time, with older patents assigned updated classifications as the system changes, so that searches of the patent record remain feasible. In contrast, the words used to describe common features may change with legal and aesthetic fashion but are not retroactively updated as the nomenclature changes.

There are more than 450 classes and more than 150,000 subclasses in the USPCS. To take two examples, class 014 corresponds to "bridges," and class 706 corresponds to "data processing (artificial intelligence)." A complete list of the current classes can be found on the USPTO website.[5] The subclasses are nested within each class and correspond to more fine-grained technological characteristics. For example, subclass 014/8 corresponds to "bridge; truss; arrangement; cantilever; suspension," while the subclass 706/29 corresponds to "data processing (artificial intelligence); neural network; structure; architecture; lattice."

Simply using the technology subclasses as components to be combined is problematic because the categories differ in their level of specificity. For example, consider three subclasses, that all belong to class 706, "data processing (artificial intelligence)":

- 706/29: Data processing (artificial intelligence); neural network; structure; architecture; lattice.
- 706/15: Data processing (artificial intelligence); neural network.
- 706/45: Data processing (artificial intelligence); knowledge processing system.

Classes 706/29 and 706/15 are both associated with neural networks, but at different levels of specificity, while 706/45 is not associated with neural networks at all. Without looking

[5] http://www.uspto.gov/web/patents/classification/selectnumwithtitle.htm

at the USPC index, it is impossible to know there is a relationship between some of the subclasses, but not others.

The uppermost subclass is called a mainline subclass, hereinafter "**mainline**." For example, the subclasses "bridge; truss," and "data processing (artificial intelligence); neural network," are both mainlines. The subclass nested one level down is said to be "one indent" in from the mainline. Within these one-indent subclasses will be still further subclasses, called two indent subclasses, and so on. Every subclass can be mapped to a mainline, but not every subclass can be mapped to one-indent or lower subclasses. Therefore, I use technology mainlines as my primary components of combination. This identifies a set comprising 13,517 components, designed to be exhaustive and nonoverlapping.

2.2. Assigning Each Patent A Combination of Mainlines

I observe the subclasses assigned to every patent[6] and for the reasons discussed above, I next collapse each technology subclass down to the mainline to which it belongs. For example, any instance of subclass 706/29, discussed above, is recoded as the mainline 706/15, since subclass 706/29 is a more specific description of the broader technology type described by mainline 706/15.

Out of 91.3 million possible mainline pairs, 1.75 million pairs are actually assigned to at least one patent over the period 1926-2009. Viewed through a combinatorial innovation lens, of the 91.3 million possible pairs, researchers have discovered how to usefully combine only 2%. Over the same period, the mean number of patents each pair belongs to over the entire period is

[6] U.S. Patent and Trademark Office (2014c). Technology classifications can be downloaded for free from http://patents.reedtech.com/classdata.php. I downloaded it in August 2014.

10.1, but the distribution is highly skewed: 51.2% of observed pairs are only ever assigned to one patent, but 46.1% of all pair assignments belong to 1% of pairs.

2.3. Dates

Patents are sequentially numbered as they are granted, so that the year any patent is granted can be inferred from the patent number.[7] Once a patent is granted, the document becomes publicly available, and its content is disclosed. I assume the information in a patent is known to other researchers beginning in the year the patent is granted.

I use the year of a patent's application to denote the year researchers develop an idea. This information is not available for all patents, but Kogan et al. (2015) extracts patent application years for every US patent from 1926-2009. Thus, although I use patent data from 1836 to construct measures of researcher knowledge, I only examine patenting behavior for the period 1926-2009. There were 6.0 million patents granted during this period.

3. Methodology

3.1. Sample

To test hypotheses 1-4, I would like to estimate equations (11) and (12), which predict the expected number of patent applications (per year) that are assigned a particular set mainlines, as a function of three explanatory variables of interest: (1) the number of times each pair of mainlines in the set has been assigned to other patents (a_{jt}), (2) the number of times the complete set has been assigned to a patent (m_{it}), and (3) the number of years the set has been available (*age*). A straightforward way to achieve this is to compute these variables for every combination of mainlines in the dataset and run a count-model regression.

[7] This can be inferred from US Patent and Trademark Office (2014a).

This straightforward strategy is computationally infeasible. With 13,517 mainlines in my dataset, there are 4.1×10^{11} unique sets of three mainlines, each of which has annual observations for the period 1926-2009, totaling more than a trillion data points. Adding in sets with four mainlines would dramatically expand the set to be searched. To obtain a more manageable dataset I restrict my attention to sets of three mainlines that are assigned to a patent at least once during the period 1926-2009. I draw a sample encompassing yearly observations on 10,000 randomly selected mainline-sets (used at least once). This gives me an unbalanced panel of 800,576 mainline-set/year observations.

Because I am restricting my attention to mainline-sets used at least once, my results are conditional and do not apply to a randomly selected set of three mainlines. Instead, they apply to a set that is assigned to a patent between 1926 and 2009. Estimating an unconditional model is made very difficult by the extreme rarity of a set of three mainlines actually being assigned to a patent. From the 4.1×10^{11} possible sets of three mainlines, only 495,369 are ever actually used at any point (less than 1 in 800,000). In contrast, once I select a mainline-set used at some point, it is actually assigned to a patent in 2.2% of years. Although the empirical exercise is restricted to this conditional dataset, in the next section I provide summary statistics for a complementary sample of 10,000 mainline-sets never assigned to any patent.

3.2. Measures

Measures of interest are described in Table 1.

Table 1. Empirical Measures

Variable	Name	Description	Intuition
y_{it}	Application Count	The number of patent applications in year t assigned mainline-set i	Dependent variable in some specifications.
$1(y_{it} > 0)$	Application Dummy	Dummy variable equal to 1 when $y_{it} > 0$.	Dependent variable in some specifications.
age_{it}	Age	The minimum number of years since a mainline in set i was first assigned to a patent.	Proxy for unobserved information that accumulates over time.
m_{it}	Mainline-Set Count	The cumulative number of patents granted up through the current period and assigned the set i (and only this set).	Proxies for the fishing out of feasible ideas.
a_{jt}	Pair Count	The cumulative number of patents granted up through the current period and assigned mainline-pair j (possibly in addition to other mainlines).	Input into my measure of affinity.
t	Time	A time-trend rescaled to 1926=0.	Used as a control.

Observations lie in the interval 1926-2009, but are constructed from data stretching back to 1836. These measures are best expressed with an example. Consider the following set of three mainlines:

- 123/319: Internal Combustion Engine; Engine Speed Regulator
- 477/34: Interrelated Power Deliver Controls, Including Engine Control; Transmission Control
- 701/1: Data Processing: Vehicles, Navigation, and Relative Location; Vehicle Control, Guidance, Operation, or Indication

In 1998, 4 patents were applied for that were assigned these three mainlines, so that $1(y_{it} > 0) = 1$ and $y_{it} = 4$. The following year, no patent applications using these three mainlines occurred, so that $1(y_{it} > 0) = y_{it} = 0$.

Mainline 123/319 was first assigned to a patent in 1860, mainline 477/34 in 1887, and 701/1 in 1923. This last example is an illustration of how the patent office updates technology classifications over time: 701/1 was first assigned to patent 1,459,106 – "Gasoline-consumption indicator for motor vehicles" which was granted in 1923. I assume that since 701/1 was first assigned in 1923, it was only feasible to combine these three technologies beginning in that year, and the *age* in 1998 is 75 years. Between 1923 and 1998, 5 other patents had already been granted that were assigned the exact same set of mainlines, so that $m_{it} = 5$.

By 1998, each of the pairs of mainlines had been used a large number of times. As of 1998, 178 patents had been granted that were assigned mainlines 123/319 and 477/34 (although not necessarily just these two mainlines). In the same year, mainlines 123/319 and 701/1 had been assigned together to patents 320 times and mainlines 477/34 and 701/1 had been jointly assigned 808 times. This data is used to estimate the probability of patentability for this mainline-set as follows:

Perfect substitutes: $$\sum_{j \in i} A_{jt} = \beta_1 (178 + 320 + 808) + \beta_2 \left(178^2 + 320^2 + 808^2\right)$$

Perfect complements: $$\min_{j \in i} \left(A_{jt}\right) = \beta_1 \times 178 + \beta_2 \times 178^2$$

Table 2 presents some summary statistics for my dataset.

Table 2. Regression Data Summary Statistics

	Min	Median	Mean	Max	St. Dev.
$1(y_{it} > 0)$	0	0	0.022	1	0.146
$y_{it} \mid y_{it} > 0$	1	1	1.265	51	1.658
age_{it}	0	89	86.97	173	37.80
m_{it}	0	0	0.935	647	5.115
a_{jt}	0	13	121.5	19,230	515.3
$\sum_{j\in i} a_{jt}$	0	108	364.4	23,010	930.9
$\min_{j\in i}\left(a_{jt}\right)$	0	2	15.61	4,124	61.66
t	1926	1969	1969	2009	24.06

As noted earlier, in most years, no patent applications are made that are assigned a given set, so that $1(y_{it} > 0) = 0$ in most cases. Conditional on $y_{it} > 0$, the average value of y_{it} is a little over 1, with a maximum of 51.

The median age_{it} of a given set of three mainlines is slightly under 90 years: most mainlines in my dataset have been available as a combination for a long time. For comparison, if each set was available in 1836, then the mean age for observations in 1926-2009 would be 131.5. The mean value of Mainline-Set Count (m_{it}) is just under 1, indicating most are used once and never again. Turning to data on pairs, we see pairs of mainlines are used together much more commonly. The minimum Pair Count (a_{jt}) in a set of three still has a mean value of 15.6.

It is important to note that this data is not representative of a randomly chosen set of three mainlines. Rather, it reflects the characteristics of mainlines eventually assigned to one patent over the period 1926-2009. For comparison, Table 3 provides statistics on a random set of 10,000 mainline-sets that are not assigned to a patent over the same period. These sets of mainlines make up the vast majority of possible combinations. Table 3 does not include data on y_{it} or m_{it} because these are all zero when restricting attention to mainline-sets that were never assigned to a patent.

Table 3. Summary Stats on a Sample of Unassigned Mainline Pairs

	Min	Median	Mean	Max	St. Dev.
age_{it}	0	52	55.60	171	34.82
a_{jt}	0	0	0.098	684	3.210
$\sum_{j \in i} a_{jt}$	0	0	0.293	684	5.582
$\min_{j \in i}\left(a_{jt}\right)$	0	0	0.001	6	0.049
t	1926	1974	1972	2009	23.36

The first thing to note is that unused sets of three mainlines tend to have a much lower *age* than those that are used at some point. The biggest difference, however, is in the number of times pairs making up a mainline-set are used. Note that $\sum_{j \in i} a_{jt}$ has a median of 0 and a mean of 0.293 for mainline-sets that are never used, compared to 108 and 364.4 for sets that are used. Sets of mainlines that are eventually used tend to have far more history of using their components together.

Tables 4 and 5 presents evidence on the correlation across these measures.

Table 4. Regression Data Correlations, all data

	$1(y_{it} > 0)$	age_{it}	m_{it}	$\sum_{j \in i} a_{jt}$	$\min_{j \in i}\left(a_{jt}\right)$
age_{it}	0.017				
m_{it}	0.135	0.110			
$\sum_{j \in i} a_{jt}$	0.084	0.218	0.250		
$\min_{j \in i}\left(a_{jt}\right)$	0.138	0.175	0.587	0.453	
t	0.050	0.538	0.115	0.265	0.172

Table 5. Regression Data Correlations, all data, $y_{it} > 0$

	$y_{it} \mid y_{it} > 0$	age_{it}	m_{it}	$\sum_{j\in i} a_{jt}$	$\min_{j\in i}\left(a_{jt}\right)$
age_{it}	0.009				
m_{it}	0.480	0.103			
$\sum_{j\in i} a_{jt}$	0.283	0.217	0.445		
$\min_{j\in i}\left(a_{jt}\right)$	0.414	0.176	0.736	0.670	
t	0.079	0.298	0.075	0.244	0.143

In all cases, measures are positively correlated, ranging from a minimum of 0.009 (between age and y_{it} when I restrict attention to observations with $y_{it} > 0$) to a maximum of 0.736 (between m_{it} and $\min_{j\in i}\left(a_{jt}\right)$ when I restrict attention to observations with $y_{it} > 0$). It is notable that in a simple correlation framework, $1(y_{it} > 0)$ and $y_{it} \mid y_{it} > 0$ are more highly correlated with $\min_{j\in i}\left(a_{jt}\right)$ than $\sum_{j\in i} a_{jt}$, suggesting the perfect complements framework may be a better fit than the perfect substitutes framework.

3.3. Estimation

Because it is relatively rare that a mainline-set is assigned to a patent in a given year, I use a two-stage estimation model:

$$E\left[y_{it}\right] = E\left[y_{it} \mid y_{it} > 0\right] \times \Pr\left(y_{it} > 0\right) \tag{13}$$

To estimate $\Pr\left(y_{it} > 0\right)$, my baseline model is a logit model. I include as explanatory variables the arguments of equations (11) and (12). I also include a time trend in the baseline model. In the baseline, I estimate clustered standard errors by resampling with replacement on mainline-sets and re-estimating coefficients. In some specifications I also use the Chamberlain estimator to strip out fixed effects from each mainline set i. A potential source of bias is variation in the propensity to innovate or patent over time. To address the potential for non-linear variation in the underlying propensity to innovate and patent, in some specifications I substitute time fixed effects for a linear time trend.

To estimate $E\left[y_{it} \mid y_{it} > 0\right]$ I run count models using either a poisson or negative binomial model, truncated below 1. In both cases, I include only observations where $y_{it} > 0$. This dramatically reduces the number of observations to just 2.2% of the original, as indicated in Table 1. I estimate clustered standard errors by resampling with replacement on mainline-sets and re-estimating coefficients.

4. Results

My results are presented in two tables, one of which uses the Perfect Substitutes framework (Table 6 and equation (11)) and the other uses the Perfect Complements framework (Table 7 and equation (12)).

Hypotheses 1, 2, and the auxiliary hypothesis are supported by these results. In all estimated specifications, $\beta_1 > 0$ and $\beta_2 < 0$ (though I cannot reject the null that $\beta_2 = 0$ in some specifications). In Table 6, the model with the minimum value of $-\beta_1/2\beta_2$ is column 4, where $-\beta_1/2\beta_2 = 4{,}551$. In my dataset, 99.74% of all observations of a_{jt} lie below 4,551, so that their marginal contribution is positive, consistent with hypothesis 1. In Table 7, the model with the minimum value of $-\beta_1/2\beta_2$ is column 1, with $-\beta_1/2\beta_2 = 1{,}259$. In my dataset, 99.97% of observations of $\min_{j\in i}\left(a_{jt}\right)$ lie below this range.

Table 6. Regression Results; Perfect Substitutes Framework

	Dependent Variable				
	$\Pr(y_{it} > 0)$	$\Pr(y_{it} > 0)$	$\Pr(y_{it} > 0)$	$E[y_{it} \mid y_{it} > 0]$	$E[y_{it} \mid y_{it} > 0]$
Time	0.014^{***}			0.018^{***}	0.019^{***}
	(0.0006)			(0.005)	(0.004)
$\sum_{j \in i} a_{jt}$	4.730^{***}	4.121^{***}	4.893^{***}	4.036^{**}	4.241^{**}
	(0.464)	(0.203)	(0.124)	(1.460)	(1.230)
$\sum_{j \in i} a_{jt}^2$	-4.426^{***}	-2.793^{***}	-4.518^{***}	-4.434	-4.354
	(0.960)	(0.192)	(0.184)	(2.588)	(2.330)
age_{it}	1.507^{***}	5.074^{***}	1.525^{***}	1.789	1.156
	(0.138)	(0.131)	(0.094)	(1.058)	(0.782)
age_{it}^2	-1.178^{***}	-2.104^{***}	-1.206^{***}	-1.642^{*}	-1.342^{**}
	(0.081)	(0.066)	(0.055)	(0.696)	(0.474)
m_{it}	6.185^{***}	-3.868^{***}	6.598^{***}	1.686	7.343^{***}
	(1.088)	(0.264)	(0.150)	(1.307)	(1.728)
m_{it}^2	-1.017	0.938^{***}	-1.073^{***}	-0.270	-1.154
	(1.155)	(0.107)	(0.038)	(0.630)	(0.901)
Constant	-4.967^{***}			-2.605^{***}	-7.770^{***}
	(0.060)			(0.393)	(0.790)
Observations	800,576	800,576	800,576	17,472	17,472
Distribution	Logit	Logit	Logit	Truncated Poisson	Truncated Neg. Bin.
Fixed Effects	None	Mainline-set	Time	None	None
Log Likelihood	-79,659.94	-62,887.78	-78,486.69	-9,333.157	-7,136.89

Notes: To make coefficients more readable, a_{jt} is measured in 10,000s, m_{it} is measured in 100s, and age_{it} is measured in centuries. Standard errors are in parentheses (clustered by bootstrapping in columns 1, 4, 5).

* = *p-value*<0.05, ** = *p-value*<0.01, *** = *p-value*<0.001

Table 7. Regression Results; Perfect Complements Framework

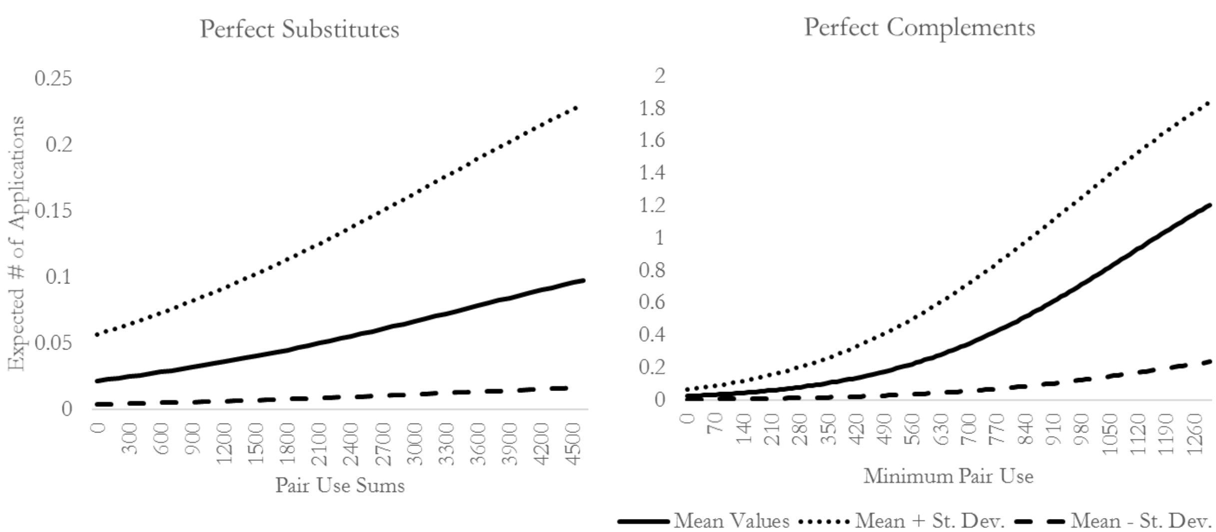
	Dependent Variable				
	$\Pr\left(y_{it}>0\right)$	$\Pr\left(y_{it}>0\right)$	$\Pr\left(y_{it}>0\right)$	$E\left[y_{it} \mid y_{it}>0\right]$	$E\left[y_{it} \mid y_{it}>0\right]$
Time	0.016^{***}			0.024^{***}	0.020^{***}
	(0.0006)			(0.004)	(0.004)
$\min_{j\in i}\left(a_{jt}\right)$	5.281^{***}	4.388^{***}	5.416^{***}	3.074^{**}	4.051^{**}
	(0.957)	(0.221)	(0.121)	(0.918)	(1.024)
$\left[\min_{j\in i}\left(a_{jt}\right)\right]^2$	-2.098	-1.172^{***}	-2.047^{***}	-0.953^{*}	-1.568
	(1.222)	(0.083)	(0.078)	(0.427)	(0.902)
age_{it}	1.599^{***}	5.443^{***}	1.611^{***}	1.395	1.120
	(0.143)	(0.131)	(0.094)	(0.868)	(0.781)
age_{it}^2	-1.239^{***}	-2.200^{***}	-1.262^{***}	-1.465^{*}	-1.347^{**}
	(0.084)	(0.066)	(0.055)	(0.589)	(0.476)
m_{it}	4.747^{***}	-5.224^{***}	5.096^{***}	1.384	6.637^{***}
	(1.090)	(0.294)	(0.163)	(1.365)	(1.776)
m_{it}^2	-0.696	1.413^{***}	-0.746^{***}	-0.213	-0.993
	(1.099)	(0.113)	(0.041)	(0.009)	(0.054)
Constant	-4.993^{***}			-2.698^{***}	-7.745^{***}
	(0.061)			(0.641)	(0.900)
Observations	800,576	800,576	800,576	17,472	17,472
Distribution	Logit	Logit	Logit	Truncated Poisson	Truncated Neg. Bin.
Fixed Effects	None	Mainline-set	Time	None	None
Log Likelihood	-79,659.94	-62,887.78	-78,3783.52	-9,169.848	-7,098.21

Notes: To make coefficients more readable, a_{jt} is measured in 1,000s, m_{it} is measured in 100s, and age_{it} is measured in centuries. Standard errors are in parentheses (clustered by bootstrapping in columns 1, 4, 5).

* = *p-value*<0.05, ** = *p-value*<0.01, *** = *p-value*<0.001

However, as Figure 1 illustrates, the actual point at which the marginal contribution of a_{jt} would turn negative is considerably higher than implied by these statistics. In Figure 1, I combine the models from column (2) and column (4) as in equation (13) to illustrate the relationship between my measure of affinity and the expected number of patent applications. In Table 4 column 2, $-\beta_1/ 2\beta_2$ = 7,377, and in Table 5 column 2, $-\beta_1/ 2\beta_2$ = 1,872, so that the true turning point obtained by multiplying these models with their counterpart in column 4 lies between 4,551 and 7,377 for $\sum_{j\in i} a_{jt}$ and 1,259-1,872 for $\min_{j\in i}\left(a_{jt}\right)$.

Figure 1. $E\left[y_{it}\right]$ as a function of $\sum_{j\in i} a_{jt}$ (left) and $\min_{j\in i}\left(a_{jt}\right)$ (right)



To compute Figure 1, I use the estimated coefficients in columns (2) and (4) of Tables 4 and 5 and compute:

$$\Pr\left(y_{it} > 0\right) = \exp\left(X'\beta\right) / \left(1 + \exp\left(X'\beta\right)\right) \tag{14}$$

$$E\left[y_{it} \mid y_{it} > 0\right] = \exp\left(X'\beta\right) / \left(1 - \exp\left(-\exp\left(X'\beta\right)\right)\right) \tag{15}$$

For equation (14), this requires estimating the mainline-set fixed effects that are stripped out by the Chamberlain estimator. Because I have slightly more than 80 observations for each mainline set, bias from the incidental parameters problem should be minimal (using Monte Carlo methods, Greene 2002 finds a bias of under 10% for samples with 20 observations per individual). To extract fixed effects, for each of 10,000 mainline sets, I solve for the α_i that maximizes the likelihood function taking all other coefficients as given:

$$L_i = \prod_i \left(\frac{\exp\left(\alpha_i + X'\theta\right)}{1 + \exp\left(\alpha_i + X'\theta\right)} \right)^{u_{it}} \left(\frac{1}{1 + \exp\left(\alpha_i + X'\theta\right)} \right)^{1-u_{it}} \tag{16}$$

where u_{it} is a dummy equal to 1 if the set i is assigned to at least one patent granted in year t.

Marginal and direct effects in non-linear models cannot be separated from other explanatory variables, and so I illustrate the impact of changing a_{jt} for three different hypothetical examples. I variously assign all other explanatory variables (including the mainline-set fixed effect) to be (1) the mean values, (2) mean values plus one standard deviation, and (3) mean values minus one standard deviation (or zero if this is negative). For Figure 1 Left (Perfect Substitutes), the horizontal axis corresponds to $\sum_{j \in i} a_{jt}$ and a there is no direct way to transform $\sum_{j \in i} a_{jt}$ into the $\sum_{j \in i} a_{jt}^2$ required of the model. However, a simple approximation:

$$\sum_{j \in i} a_{jt}^2 = 0.8752 \times \left[\sum_{j \in i} a_{jt} \right]^2 + \varepsilon \tag{17}$$

fits the data very well, with an $R^2 = 0.96$. Therefore, I approximate $\sum_{j\in i} a_{jt}^2$ with

$0.8752 \times \left[\sum_{j\in i} a_{jt}\right]^2$ purely for the sake of illustration in Figure 1. For Figure 1 right (Perfect Complements), the horizontal axis corresponds to $\min_{j\in i}\left(a_{jt}\right)$ and it is straightforward to compute the additional explanatory variable $\left[\min_{j\in i}\left(a_{jt}\right)\right]^2$.

As Figure 1 illustrates, an increase in a_{jt} is associated with an increase in $E[y_{it}]$, but the marginal effect does not really begin to decline over the illustrated range (which encompasses most of the observations). This is consistent with a learning story where, if the impact of learning ever declines, it only does so only at very high levels of a_{jt}. I defer a discussion of the magnitude of these effects until the next section.

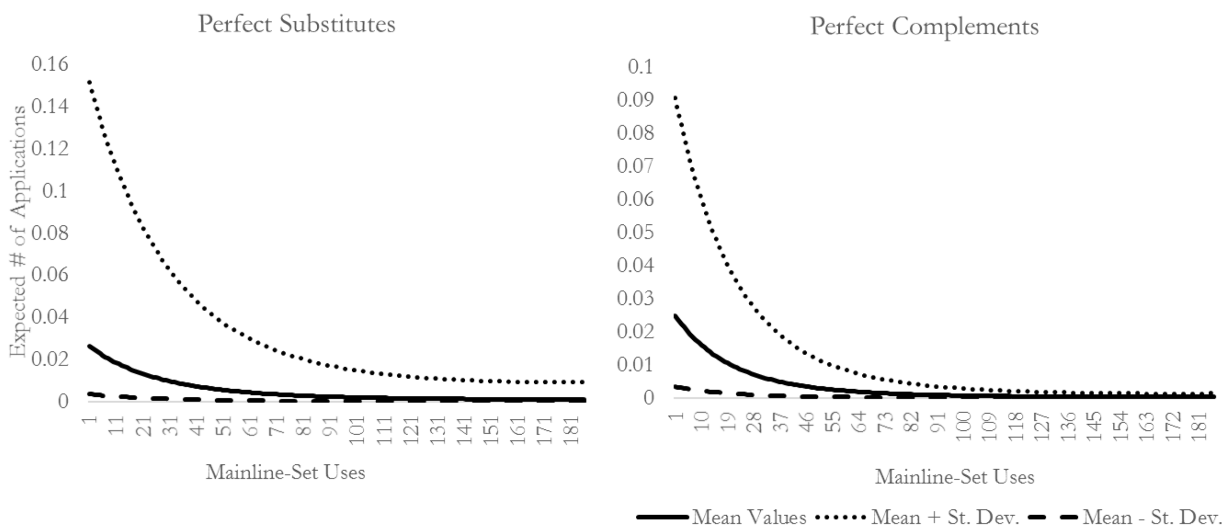
Hypothesis 3 finds more equivocal support. In columns 1 and 3-5, m_{it} is positively correlated with $E[y_{it}]$ (although I cannot reject the null that the coefficient is zero in column 4), in defiance of hypothesis 3. However, these models do not control for variation in the value of mainline-sets. If some mainline-sets are persistently more valuable (or R&D is less costly), then these mainline-sets will be attempted more often, introducing an upward bias into my estimate of ϕ_1 and ϕ_2. In column 2, when I use the Chamberlain estimator to strip out mainline-set fixed effects, the coefficients take the expected direction.

For Table 4, Column 2, $-\phi_1 / 2\phi_2 = 205.7$ while in Table 5, Column 2, $-\phi_1 / 2\phi_2 = 184.9$. In my dataset, 99.99% of observations of m_{it} lie below 184.9. So long as I include fixed effects, each patent assigned mainline-set i reduces the expected number of further such applications.

Modeling fixed effects in a truncated poisson is beyond the scope of this paper. However, as illustrated in figure 2, even if fixed effects do not upwardly bias my estimates of ϕ_1 and ϕ_2 in columns 4 and 5, my full model exhibits the expected negative relationship between m_{it} and

$E[y_{it}]$. To compute Figure 2, I use the same approach as in Figure 1; extracting the fixed effects and plotting $E[y_{it}]$ for three cases: mean values for other variables, mean less one standard deviation, and mean plus one standard deviation.

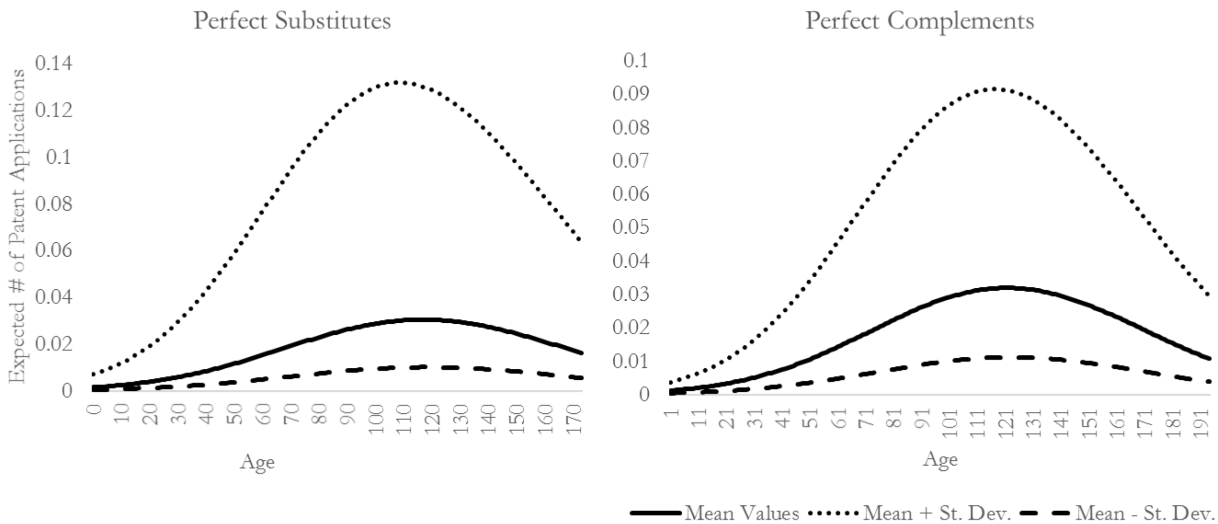
Figure 2. $E\left[y_{it}\right]$ as a function of m_{it}



Finally, hypothesis 4 is also supported. In all cases, I find $\gamma_2 < 0$, so that the marginal impact of age is declining in the long run. Unlike the other estimated coefficients, the relationship between patent applications and age is non-monotonic. Depending on the specification chosen from Tables 4 and 5, $-\alpha_1 / 2\alpha_2 \in \left[41.6, 123.7\right]$ with a mean of 68.7 years and a median of 63.5 years. In the early years after a set of mainlines becomes available, each passing year increases the expected number of patents applications making use of the mainlines. After approximately 65 years, additional years begin to subtract from the expected number of patent applications that will be made using the mainlines. In Figure 3, I combine the models from column (2) and column (4) of Tables 4 and 5, as in equation (13), to diagram the relationship between the availability age of a mainline-set and the expected number of patent applications.

For the values of the other explanatory variables, I again use mean values plus or minus one standard deviation.

Figure 3. $E\left[y_{it}\right]$ as a function of Age



Both the Perfect Substitutes and Perfect Complements frameworks do well in predicting the number of patent applications in a given year. Of the two, the Perfect Complements approach performs slightly better in terms of the log likelihood of the model, and has a much less flat relationship between the expected number of patent applications and the value of the explanatory variable. However, in columns 4 and 5, the coefficients on the perfect substitutes value have slightly better *p* values.

5. Exploring Demand Side Interpretations

Throughout the paper, I have assumed that changes in a_{jt} and m_{jt} drive $E[y_{it}]$ through the learning and fishing out channels. This assumption need not be true. An alternative interpretation of my results is that changing *demand* for various technology bundles drives my results.

To see how this might be, rewrite the definition of V_{dt} (given in equation 3) as follows:

$$V_{dt} = v_i + u_t^v + \sum_{j \in i} w_{jt} + Z(m_{it}) + e_{dt}^v \tag{18}$$

where $Z(m_{it})$ is a function of m_{it}. If $Z' < 0$, then the value of ideas falls as more patents are granted using the same mainline-set. This can occur if technologies using the same set of components compete with each other for the same market. As more competitors enter the same space, each enjoys a smaller share of consumer demand, diminishing the value of such inventions. In this case, the negative correlation between m_{it} and $E[y_{it}]$ is driven by greater competition in a given market (making it less desirable for new entrants), rather than the finite supply of distinct patentable combinations. These effects are conceptually quite similar. In either case, R&D has the impact of reducing the number of R&D projects with positive expected value, and my interpretation of the coefficient on m_{it} is not substantively changed.

The term w_{jt} in equation (18) is a time-varying error term associated with mainline-pair j. If these new error terms are correlated over time, this would suffice to generate a positive correlation between a_{jt} and y_{it} driven by shifting demand, rather than learning. If this is the case, a_{jt} is merely picking up lagged demand for certain packages of technology. The results would therefore have nothing to say about learning and spillovers.

5.1 A Model of Patent Renewal

To explore the validity of demand-side interpretations of my result, I use patent renewal data. Since 1982, US patents have been subject to a renewal fee at 4, 8, and 12 years. These fees increase at each stage, and if not paid, the patent expires. There is a long tradition[8] of using

[8] See Pakes and Schankerman (1984), Lanjouw et al. (1998), Baudry and Dumont (2006), Bessen (2008), Serrano (2010).

patent renewal data to infer the value of patents, based on the assumption that patents are only renewed if the value of an active patent exceeds the renewal fee.

Let us consider a patent that was applied for in year t, has been renewed at 4 and 8 years, and where the patent-holder must decide whether to renew at 12 years in year T at cost W_{dT}. Suppose the patent would be worth V_{dT} if it had been applied for in year T. However, because the patent is in force for fewer years, its renewal value is only a fraction δ of this value. Patents are renewed if $\delta V_{dT} \geq W_{dT}$. This condition can be written as:

$$\Pr\left(d \text{ renewed} \mid d \text{ granted}\right) = \Pr\left(\delta V_{dT} > W_{dT}\right) \tag{19}$$

Adding and subtracting V_{dt} (value at the time of application), and using (18), this can be rewritten as:

$$\begin{aligned}&\Pr\left(d \text{ renewed} \mid d \text{ granted}\right) = \\ &\Pr\left(\delta\left[V_{dt} + \left(u_T^v - u_t^v\right) + \sum_{j\in i}\left(w_{jT} - w_{jt}\right) + \left(\mathrm{Z}\left(m_{iT}\right) - \mathrm{Z}\left(m_{it}\right)\right) + \left(e_{dT}^v - e_{dt}^v\right)\right] > W_{dT}\right)\end{aligned} \tag{20}$$

Equation (20) gives the probability of renewal as a function of value at the time of application, plus changes that have occurred since application. In particular, if Z' < 0, then the probability of renewal declines when m_{it} increases *after* application. If $w_{jT} - w_{jt} > 0$, then renewal is more likely. However, if w_{jt} is positively correlated over time, then $w_{jT} - w_{jt} > 0$ means it is also more likely that $w_{jt'} > w_{it}$ for $t < t' < T$. In each of these years, ideas containing pair j will be more valued, ultimately leading to more patent grants for these ideas and a higher value of a_{jt}. Thus, if w_{jt} is positively correlated over time, the probability of renewal will be positively correlated with changes in a_{jt} that occur *after* application.

Equation (20) motivates the following reduced form regression:

$$\Pr\left(d \text{ renewed} \mid d \text{ granted}\right) = \text{logit}\left(\lambda_1 \underline{a}_{it} + \lambda_2\left(\underline{a}_{it+12} - \underline{a}_{it}\right) + \omega_1 m_{it} + \omega_2\left(m_{it+12} - m_{it}\right) + X'\theta\right) \tag{21}$$

where $\underline{a}_{it} \equiv \min_{j \in i}\left(a_{jt}\right)$ to economize on space and X includes a number of controls. Equation (21) is a reduced form model of the renewal decision, where the probability of renewal is a logit function. The variable m_{it} is a measure of mainline-set count at the time of application, and $(m_{it+12} – m_{it})$ is the change in mainline-set count between application and the renewal decision. The independent variable $\underline{a}_{it}$ is the minimum pair-count at the time of the patent application and $\left(\underline{a}_{it+12} - \underline{a}_{it}\right)$ is the changes in minimum pair count between application and the renewal decision.

If $\omega_2 < 0$, then patents are less likely to be renewed if other patents with the same set of technologies are granted in the years between application and renewal. This would be consistent with fishing out driven by falling demand for identical technology bundles. If $\lambda_2 > 0$, patents are more likely to be renewed if other patents using the same technologies are granted in the years *after* the patent application. This would suggest some of the positive correlation between a_{jt} and $E[y_{it}]$ is driven by time-varying changes in demand.

This framework also allows us to test an additional hypothesis:

Hypothesis 5 (selection effect): $\lambda_1 < 0$

Hypothesis 5 follows from equation (20), which says the probability of renewal is increasing in value at the time of application and equation (5), which says firms will only attempt R&D projects if their net expected value is positive. Equation (5) implies any idea that is attempted in spite of the low affinity of its components has either a high value of V_{dt} or a low value of K_{dt}. Patents are more likely to be renewed if V_{dt} is high, and this is more likely to obtain if $\underline{a}_{it}$ is smaller. In plain words, it is only worth initiating R&D on long-shot projects if the payoff from success is high. When these long-shots pay off, they result in patents that are more likely to be worth renewing.

5.2 Data

Data on patent renewals is available from the USPTO PatentsView website. There are 368,721 patents with three mainlines eligible for renewal fees and facing the 12-year renewal decision by 2012 (the last year I have data on a_{jt} and m_{it}). I restrict my attention to patents with US firms or individuals as the assignees, which reduces my sample to 154,774 patents. Finally, because I wish to include fixed effects at the level of mainline-sets, I further restrict my sample to mainline-sets that are used by at least two patents in this restricted data set. This leaves me 12,274 mainline-sets spread over 50,595 patents. In this sample, 77.6% of patents were renewed at year 12. James Bessen kindly provided maintenance fee data from Bessen (2008).

In this highly selected sample of patents, I find the mean minimum pair count ($\underline{a}_{it}$) and the mean mainline-set count (m_{it}) at the time of application is significantly higher than in the full data sample. The mean $\underline{a}_{it}$ is 382 compared to 15 for the full dataset, and the mean m_{it} is compared to 0.9 in the full dataset. These large differences are primarily attributable to the highly selected nature of the sample (only mainline-sets used multiple times between 1980-2000), and because I am looking at pair count at the time of application, instead of at every year beginning with the mainline-set's availability.

There is significant activity in the same technological space in the years between application and the renewal decision. Minimum pair count at renewal time has a mean 655 greater than at application, and a median 106 greater. And in the intervening years, the difference between the mainline-set count at application and renewal has a mean of 85 and a median of 12.

More information about this dataset and some of the controls used in the regressions are available in the appendix.

5.3 Results

Results are displayed in Table 8.

Table 8. Patent Renewal Logit Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
log(Fee)	-1.912^{***}	-1.915^{***}	-1.905^{***}	-0.483^{***}	-0.492^{***}	-0.484^{***}
	(0.229)	(0.229)	(0.230)	(0.136)	(0.136)	(0.136)
Large Entity	1.888^{***}	1.882^{***}	1.882^{***}	0.471^{***}	0.477^{***}	0.471^{***}
	(0.162)	(0.162)	(0.162)	(0.097)	(0.097)	(0.097)
Application Year	0.052^{***}	0.052^{***}	0.052^{***}	0.011^{***}	0.010^{***}	0.010^{***}
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
1(App. Year > 1995)	-0.041	-0.039	-0.045	-0.040^{**}	-0.040^{**}	-0.041^{**}
	(0.034)	(0.034)	(0.034)	(0.019)	(0.019)	(0.019)
$\underline{a}_{it}$	-0.152^{***}		-0.195^{***}	-0.049^{**}		-0.084^{**}
	(0.011)		(0.021)	(0.024)		(0.040)
$\underline{a}_{it+12} - \underline{a}_{it}$	0.024^{***}		0.010	-0.037^{**}		-0.011
	(0.008)		(0.014)	(0.017)		(0.025)
m_{it}		-0.116^{***}	0.057^{***}		-0.017	0.034
		(0.011)	(0.021)		(0.018)	(0.030)
$m_{it+12} - m_{it}$		0.015^{***}	0.011		-0.035^{***}	-0.025
		(0.005)	(0.008)		(0.012)	(0.018)
Constant	14.887^{***}	14.907^{***}	14.842^{***}			
	(1.758)	(1.756)	(1.758)			
Observations	50,595	50,595	50,595	50,595	50,595	50,595
Fixed Effects	N	N	N	Y	Y	Y
Log Likelihood	-26,419	-26,466	-26,412	-76,300	-76,301	-76,298

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Note: To make coefficients more readable, a_{jt} is measured in 1,000s, m_{it} is measured in 100s

Columns 1-3 are logit regressions, with the dependent variable equal to 1 if the patent is renewed at year 12. Columns 4-6 are also logit regressions with the same dependent variable, but I use the Chamberlain estimator to eliminate mainline-set fixed effects. Because the minimum pair-count and mainline-set count are highly correlated, I present three regression models without fixed effects and three with fixed effects: minimum pair count only (columns 1 and 4), mainline-set count only (columns 2 and 5), and both counts (columns 3 and 6).

Hypothesis 5 is supported in all applicable models. The probability a patent will be renewed after 12 years is higher for more novel patents (lower minimum pair count at the time of application), consistent with the predictions of the selection effect implied in equation (5).

The coefficients on $\underline{a}_{it+12} - \underline{a}_{it}$ and $m_{it+12} - m_{it}$ are unstable and change sign or lose statistical significance when additional explanatory variables are used. In columns 1 and 2, the coefficients are positive and statistically different from zero. However, when both variables are included together in column 3, they are both positive but I cannot reject the null hypothesis that each (individually) has zero effect. The positive coefficients, however, appear to be driven by the presence of mainline-set fixed effects. As noted in section 4, if some mainline-sets are persistently more valuable, this omitted variable bias will upwardly bias the coefficients on measures of a_{jt} and m_{it}. In this case, mainline-sets that are more valuable get renewed more often *and* have more patents using the mainlines granted in the years between application and renewal.

When I include mainline-set fixed effects, as in columns 4-6, the coefficients on $\underline{a}_{it+12} - \underline{a}_{it}$ and $m_{it+12} - m_{it}$ change signs. The negative coefficient on $\underline{a}_{it+12} - \underline{a}_{it}$ in column (4) strongly rejects the alternative hypothesis that the positive correlation between a_{jt} and $E[y_{it}]$ is driven merely by time-varying demand for different technological bundles. The negative coefficient on $m_{it+12} - m_{it}$ however, does lend support to the demand side fishing out effect discussed in 7.2. One possible interpretation of the negative coefficient on $\underline{a}_{it+12} - \underline{a}_{it}$ in column (4) is that it is also picking up a more diffuse set of demand side fishing out, whereby certain bundled technology features crowd each other out. If this is the case, then my estimates on the learning effect may be mixing the positive learning effect and a negative demand side fishing out effect. If this is the case, my results understate the extent of learning. In any event, when both $\underline{a}_{it+12} - \underline{a}_{it}$ and $m_{it+12} - m_{it}$ are included, as in column (6), in neither case can I reject the null that the (individual) effect is actually zero. These results support my interpretation of the main results as deriving from learning and fishing out effects.

6. Discussion: Learning or Fishing Out?

Returning to my main results in section 4, while the estimated coefficients are statistically significant in the expected directions, figures 1-3 indicate marginal changes to a_{jt}, m_{it} and age have a small impact on the dependent variable $E[y_{it}]$. For example, with the exception of Figure 1 (right), the maximum illustrated value of $E[y_{it}]$ in any figure is under 0.25, even when explanatory variables take on very large values. However, it must also be recalled that (1) the unconditional value of $E[y_{it}] = 0.028$, (2) the fishing out effect is invariably bound up with the learning effect, making it difficult to interpret in isolation, and (3) learning has positive spillovers to other ideas. I take each of these complications in turn using as my preferred benchmark the Perfect Complements framework obtained by combining Table 5 (column 2) with Table 5 (column 4).

To begin, while the marginal impact of patenting is small in absolute terms, the magnitude of the effect is more substantive in relative terms. To measure the empirical import of affinity in isolation, for every observation I increase $\min_{j \in i}\left(a_{jt}\right)$ by one standard deviation (61.7) and compare the new $E[y_{it}]$ to the old. If I subtract the new value from the old, the mean increase in $E[y_{it}]$ is 0.008, but if I divide the new value by the old, the mean proportional increase is 31.9%. Conversely, increasing m_{it} by one standard deviation (5.1) reduces $E[y_{it}]$ by an average of 0.005 or 22.1%.

Second, whenever a viable idea is discovered, there are two opposing effects. The successful combination of technological components has a positive learning effect, because it raises the affinity between components. It also has a negative effect, because it uses up one possible combination of technological components. Which effect dominates depends on how much firms already know about the affinity of the pairs in question. Because this paper separately identifies the learning and fishing out effects, the reduced form model estimates shed some light on when each effect dominates.

First, consider these two effects on a particular set of mainlines. Whenever a mainline set is patented, it increases both m_{it} and $\min_{j \in i}\left(a_{jt}\right)$ by one. When using the Perfect Complements framework, the fishing out effect dominates the learning effect at the level of a mainline-set if the following condition holds:

$$\beta_1\left(\underline{a}_{it}+1\right)+\beta_2\left(\underline{a}_{it}+1\right)^2+\phi_1\left(m_{it}+1\right)+\phi_2\left(m_{it}+1\right)^2<\beta_1\underline{a}_{it}+\beta_2\underline{a}_{it}^2+\phi_1 m_{it}+\phi_2 m_{it}^2 \qquad (22)$$

Which can be expressed as:

$$-\frac{1}{2\beta_2}\left\{\beta_1+\beta_2+\phi_1+\phi_2+2\phi_2 m_{it}\right\}<\underline{a}_{it} \qquad (23)$$

Using the coefficients from Table 7, column 2 (the fixed effect model) and converting into consistent units (remembering that I measured a_{jt} in 1,000s and m_{it} in 100s to facilitate display in the table), the condition under which fishing out dominates learning is:

$$120.6m_{it} - 20,355 < \underline{a}_{it} \tag{24}$$

This condition is most likely to fail when m_{it} is "large" and $\underline{a}_{it}$ is "small." Note the minimum value of $\underline{a}_{it}$ is m_{it} by definition, implying the fishing out effect always dominates for $m_{it} < 170.2$ (corresponding to nearly all observations). Taking the reduced form model literally, it is possible for the learning effect to dominate above $m_{it} = 170.2$ if $\underline{a}_{it}$ is sufficiently small (implying there are few other patents using the pair), because by this point the negative effect of fishing out has been dissipated. However, the underlying theory justifying the reduced form model implies the fishing out effect only disappears when there are no ideas left to try, which would mean there are no untried ideas left to apply better information about affinity towards.

Thus, it would seem the fishing out effect always dominates the learning effect, when we restrict our attention to a single set of mainlines. Every time that set of mainlines is combined, the expected number of patents that will use this set in the future declines. This does not imply the fishing out effect dominates the learning effect on the whole though, because every successful combination has positive spillovers for a large number of other ideas.

To investigate the full magnitude of the learning effect, observe that across all observation, if we increase $\min_{j \in i}\left(a_{jt}\right)$ by 1, the average increase in $E[y_{it}]$ is 1.1x10^{-4} and if we increase m_{it} by 1, the average decrease in $E[y_{it}]$ is 1.1x10^{-3}.

Restricting attention to the set of 495,369 mainline-sets that are combined at some point in 1926-2009, a typical set has a pair of mainlines in common with 50 other mainline-sets. If we assume there is a 1 in 3 chance that the shared mainline is the pair with the minimum a_{jt} (while this is probably an overestimate, neither is it true that pairs are perfect complements and only the minimum a_{jt} contributes), then each time a mainline-set is patented, it increases the expected number of patent applications for other combinations by $0.00011 \times 50 / 3 = 1.8\text{x}10^{-3}$ per year. Thus, the total learning effect ($1.8\text{x}10^{-3}$ additional patents per year) exceeds the fishing out effect ($1.1\text{x}10^{-3}$ fewer patents per year).

Because these effects largely cancel out, the net impact is small. Over the course of a century, the net impact is that granting a patent with three mainlines is correlated with an additional 0.07 patents with three mainlines. That said, two caveats are in order. First, these results apply only to the subset of patents with three mainlines, and therefore must represent an underestimate of the total impact. Second, the impact of learning should grow over time as the set of technological components expands. As Weitzman (1998) anticipated, there is a tipping point in combinatorial growth. When the set of components is too small (under 10 spillovers per patent in my preferred model) the fishing out effect exceeds the learning effect and every innovation, on net, reduces the extent of future innovation. Once an economy is past this threshold, there are enough applications for new knowledge so that the learning effect exceeds the fishing out effect going forward.

7. Conclusions

I describe a model where ideas are created by combining pre-existing technological components. The probability a set of components yields a viable idea is a function of the "affinities" its components have for each other. The affinity between a pair of components measures how well researchers know how to combine them. The more a pair of components has been successfully combined, the higher the affinity of the pair.

Thus, there is a positive spillover from the discovery of a successful combination. These discoveries provide a new example of how to usefully combine components. This raises the affinity these components have for each other. In turn, this raises the probability ideas with some of the same components are viable. At the same time, the discovery of a successful combination leaves one fewer idea to be discovered. Throughout the paper, the first effect is called the learning effect and the second the fishing out effect.

This model lends itself to empirical application. I use the US patent classification system to define a set of 13,517 technological components that patents reuse and recycle. My dataset is a panel of 10,000 sets of three technological components. For each year and each set, I observe the number of patent applications using these components, as well as additional explanatory variables. Patenting increases the future number of similar patents (those that share two of the three technological components), but decreases the number of identical patents (those that share all three technological components).

This finding is consistent with my model's predictions. Predictions about the curvature of the relationship between patent applications, the age of the set, and my proxy for affinity are also supported. In an extension, I also show firms are more likely to renew patents with a low affinity at the time of their application. Because firms must pay to renew patents, this suggests these patents are more valuable. This is consistent with my model: researchers only attempt ideas unlikely to be viable (those with low affinity) if they are very valuable when viable.

I also use patent renewal data to rule out an alternative interpretation of my main finding. Changing demand for certain bundles of technology can also generate some of my results. For example, if demand for patents with technologies x and y increases, we will see an increase in patent applications with x and y, as well as an increase in my proxy for the affinity of x and y. Demand, not learning, drives any positive correlation between the two. If this is the case though, the decision to renew a patent will be positively correlated with the number of similar patents granted after patent application. After controlling for set fixed effects, either there is no correlation, or it is negative.

Limitations remain. The model contains several simplifying assumptions: firms are myopic, only engage in one project, and make a binary choice to either initiate research or not. The set of technological components is fixed and exogenous, rather than endogenously determined. On the empirical side, my data is also limited. Technology subclasses are a very coarse proxy encompassing a variety of distinct technologies. Moreover, to identify the impact of the learning and fishing out effects, I restricted my analysis to patents with three mainlines. These may not be representative. This selection problem is even more acute with my renewal data, where I rely on sets of technological components with at least two patents renewed. These constitute a very exclusive set.

Future research could address these limitations. In particular, locating more precise measures of technological components (i.e., words in patents, citations, or subclasses at a more disaggregated level) could clarify if the explanatory power of these regressions is attenuated by measurement error. This model also suggests applications for measuring spillovers and knowledge diffusion. Furthermore, firm level data could be used to see if R&D is more productive when this model predicts. It would also be possible to extend this analysis to non-patent domains where appropriate proxies for the elements of combination are available.

Turning to policy implications, this paper has some tentative implications for R&D policy. Because the learning effect is a positive externality, R&D will be under-provided by private actors. R&D subsidies can equalize the private and social returns to R&D, but the size of the subsidy needs to reflect the size of the externality. This paper has some guidance for measuring the size of potential learning externalities. I find a perfect complements framework fits better than the perfect substitutes framework. This implies the expected viability of an idea is most impacted when the lowest-affinity pair is strengthened. Put another way, spillovers are more potent when they touch ideas with high affinity among the unshared components. At the same time, Figure 2 indicates the link between a_{jt} and $E[y_{it}]$ is convex for low values of a_{jt} (and concave for high values). Thus, the marginal impact of increasing a_{jt} is itself increasing (up to the inflection point). Together, this suggests the ideas with the greatest spillover potential are those that strengthen the affinity of the "weakest link" of a large number of other ideas, but where the affinity of the "weakest link" is not too low itself.

To return to the question that motivated this paper, I find both innovation optimists and pessimists have grounds for their beliefs. For the optimists, I find evidence that every discovery gives us knowledge applicable to new contexts. This knowledge enables new discoveries that, in turn, give us knowledge applicable in still other contexts. This process can repeat, and it is possible for innovation to become self-propelling. For the pessimists, I find evidence that ideas can be "used up" like a natural resource. Every time a patent is discovered, fewer patents with the same technologies arrive in subsequent years.

Within a narrow technical domain, the fishing out effect is strongest. When some combination *xyz* is patented, researchers do not learn enough about how to combine *x*, *y*, and *z* to overcome the negative effect of fishing out one of the *xyz* configurations. The learning effect, however, spills over to many adjacent technical domains. On average, these spillovers are so numerous that the learning effect exceeds the fishing out effect. This is true, even within my restricted dataset of three-mainline patents. Noting all the caveats mentioned above, this paper is thus more consistent with the position of the optimists than the pessimists. Every new discovery opens (slightly) more doors than it closes.

Acknowledgments:

This work is an extension of my dissertation, and I am grateful for many helpful comments from my advisor, GianCarlo Moschini. This work also benefitted from the comments of an anonymous reviewer. Lastly, James Bessen kindly provided data on patent maintenance fees. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The views expressed are those of the author and should not be attributed to the Economic Research Service or USDA.

References

Allen, Robert C. 2009. *The British Industrial Revolution in Global Perspective*. Cambridge, UK: Cambridge University Press.

Arthur, W. Brian. 2009. *The Nature of Technology: What It Is and How It Evolves.* New York: Free Press.

Akcigit, Ufuk, William R. Kerr, and Tom Nicholas. (2013). The Mechanics of Endogenous Innovation and Growth: Evidence from Historial U.S. Patents. Working Paper.

Auerswald, Philip, Stuart Kauffman, Jose Lobo, and Karl Shell. 2000. "The Production Recipes Approach to Modeling Technological Innovation: An Application to Learning By Doing." *Journal of Economic Dynamics and Control 24* 389-450.

Baudry, Marc, and Béatrice Dumont. 2006. Patent Renewals as Options: Improving the Mechanism for Weeding Out Lousy Patents. *Review of Industrial Organization* 28(1): 41-62.

Bessen, James. 2008. The value of U.S. patents by owner and patent characteristics. *Research Policy* 37(5): 932-945.

Bostrom, Nick. 2016. *Superintelligence: Paths, Dangers, Strategies*. Oxford, UK: Oxford University Press.

Brynjolfsson, Erik, and Andrew McAfee. 2014. *The Second Machine Age: Work, Progress and Prosperity in a Time of Brilliant Technologies.* New York; W.W. Norton & Company, Inc.

Cowen, Tyler. 2011. *The Great Stagnation: How America Ate All the Low-Hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better*. New York: Dutton Publishing.

Dartnell, Lewis. (2014). *The Knowledge: How to Rebuild Our World From Scratch*. New York, NY: Penguin Press.

Doudna, Jennifer A., and Samuel H. Sternberg. 2017. *A Crack in Creation: Gene Editing and the Unthinkable Power to Control Evolution*. New York: Houghton Mifflin Harcourt.

Feinstein, Jonathan. 2011. Optimal Learning Patterns for Creativity Generation in a Field. *American Economic Review Papers and Proceedings*: 1-5.

Fleming, Lee. 2001. "Recombinant Uncertainty in Technological Search." *Management Science 47(1)* 117-132.

Ghiglino, Christian. 2012. Random Walk to Innovation: Why Productivity Follows a Power Law. *Journal of Economic Theory* 147: 713-737.

Gordon, Robert. 2016. *The Rise and Fall of American Growth: The U.S. Standard of Living Since the Civil War*. Princeton, NJ: Princeton University Press.

Greene, William. 2002. The Behavior of the Fixed Effects Estimator in Nonlinear Models. NYU Working Paper No. EC-02-05. Available at SSRN: https://ssrn.com/abstract=1292651

Hall, Bronwyn, Adam Jaffe, and Michael Trajtenberg. 2001. The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. NBER Working Paper 8498.

Jovanovic, Boyan and Rafael Rob. 1990. "Long Waves and Short Waves: Growth Through Intensive and Extensive Search." *Econometrica* 1391-1409.

Kaplan, Sarah, and Keyvan Vakili. 2015. The Double-Edged Sword of Recombination in Breakthrough Innovation. *Strategic Management Journal* 36(10): 1435-1457.

Kauffman, Stuart, Jose Lobo, and William G. Macready. 2000. "Optimal Search on a Technology Landscape." *Journal of Economic Behaviour and Organization 43* 141-166.

Keijl, S., V.A. Gilsing, J. Knoben, and G. Duysters. 2016. "The Two Faces of Inventions: The Relationship Between Recombination and Impact in Pharmaceutical Biotechnology." Research Policy 45(5): 1061-1074.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. (2015). Technological Innovation, Resource Allocation and Growth. http://ssrn.com/abstract=2193068.

Lanjouw, Jean O., Ariel Pakes, and Jonathan Putnam. 1998. How to Count Patents and Value Intellectual Property: The Uses of Patent Renewal and Application Data. *The Journal of Industrial Economics* 46(4): 405-432.

Meisenzahl, Ralf, and Joel Mokyr. 2011. The Rate and Direction of Invention in the British Industrial Revolution: Incentives and Institutions. *NBER Working Paper* 16993.

Mokyr, Joel. 1990. *The Lever of Riches: Technological Creativity and Economic Progress*. Oxford, UK: Oxford University Press.

Nemet, Gregory F. 2012. "Inter-technology Knowledge Spillovers for Energy Technologies." *Energy Economics 34* 1259-1270.

Nemet, Gregory F., and Evan Johnson. 2012. "Do Important Inventions Benefit from Knowledge Originating in Other Technological Domains?" *Research Policy 41* 190-200.

Olsson, Ola. 2005. Technological Opportunity and Growth. *Journal of Economic Growth* 10(1): 35-57.

Olsson, Ola, and Bruno S. Frey. 2002. Entrepreneurship as Recombinant Growth. *Small Business Economics* 19(2): 69-80.

Pakes, Ariel, and Mark Schankerman. 1984. The rate of obsolescence of patents, research gestation lags, and the private rate of return to research resources. In *R&D, Patents and Productivity*, ed. Zvi Griliches. Chicago: University of Chicago Press for the NBER.

Poincaré, Henri. 1913. *The Foundation of Science: Science, Hypothesis, The Value of Science, Science and Method*. New York: The Science Press.

Sawyer, R. Keith. 2012. *Explaining Creativity: The Science of Human Innovation (second edition)*. Oxford, UK: Oxford University Press.

Schilling, Melissa A., and Elad Green. 2011. "Recombinant Search and Breakthrough Idea Generation: An Analysis of High Impact Papers in the Social Sciences." *Research Policy 40* 1321-1331.

Schoenmakers, Wilfred and Geert Duysters. 2010. "The Technological Origins of Radical Inventions." *Research Policy 39* 1051-1059.

Serrano, Carlos J. 2010. The dynamics of the transfer and renewal of patents. *The RAND Journal of Economics* 41(4): 686-708.

Simonton, Dean Keith. 2004. *Creativity in Science: Chance, Logic, Genius, and Zeitgeist.* Cambridge, UK: Cambridge University Press.

United States Patent and Trademark Office. 2012. "Overview of the U.S. Patent Classification System (USPC)." *The United States Patent and Trademark Office.* 12. Accessed 10 15, 2014. http://www.uspto.gov/patents/resources/classification/overview.pdf.

—. 2014a. *Table of Issue Years and Patent Numbers, for Selected Document Types Issued Since 1836.* March 26. Accessed October 16, 2014. http://www.uspto.gov/web/offices/ac/ido/oeip/taf/issuyear.htm.

—. 2014b. "U.S. Manual of Classification File (CTAF)." *USPTO Data Sets.* Accessed August 2014. http://patents.reedtech.com/classdata.php.

—. 2014c. "U.S. Patent Grant Master Classification File (MCF)." *USPTO Data Sets.* Accessed August 2014. http://patents.reedtech.com/classdata.php.

Usher, Abbott Payson. 1929. *A History of Mechanical Inventions*. New York: McGraw-Hill.

Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Ben Jones. 2013. Atypical Combinations and Scientific Impact. *Science* 342(6157): 468-472.

Vance, Ashlee. 2015. *Elon Musk: Tesla, SpaceX, and the Quest for a Fantastic Future*. New York: HarperCollins Publishing.

Weitzman, Martin L. 1998. "Recombinant Growth." *Quarterly Journal of Economics 2* 331-360.

Wootton, David. 2015. *The Invention of Science: A New History of the Scientific Revolution*. New York: HarperCollins Publishing.

Appendix: Patent Renewal Data

Data Description

As discussed in the paper, data on patent renewals is available from the USPTO PatentsView website. After restricting my attention to patents (1) eligible for renewal fees, (2) facing the 12 year renewal decision prior to 2013, (3) assigned to US firms or individuals, and (4) assigned a set of three mainlines that appears more than once in the dataset, I have 12,274 mainline-sets spread over 50,595 patents. Table A1 provides summary statistics on this subset of patents.

Table A1. Patent Renewals Data Summary

Statistic	Min	Median	Mean	Max	St. Dev.
Renewed	0	1	0.776	1	0.417
Fee (2016 $)	$1,984	$4,319	$4,058	$4,953	$768
Large Entity	0	1	0.865	1	0.342
Application Year	1980	1993	1992	2000	4.439
$\underline{a}_{it}$	0	87	383	10,560	1,028
$\underline{a}_{it+12} - \underline{a}_{it}$	0	106	655	13,370	1,623
m_{it}	0	7	34.5	1,309	98.4
$m_{it+12} - m_{it}$	0	12	84.9	3,073	274

The regressions displayed in Table 8 of the main paper include a number of additional controls that were not discussed in the interest of conserving space. Here I briefly discuss these controls and their associated regression coefficients.

Fees: James Bessen kindly provided maintenance fee data from Bessen (2008) which I converted into 2016 dollars. Renewal fees changed both over time, and depending on the status of the patent-holder, with an average fee of $4,058. In every specification, the probability of renewal is negatively correlated with the log of the fee.

Large Entity: Firms classified as large entities pay fees twice as large as small entities, and Bessen (2008) finds this variable is an important predictor of the renewal decision. For patents that are renewed at 12 years, I use the designated entity status as provided by the USPTO. For firms that do not renew at 12 years, I use the designated entity status at the year 8 renewal decision. By these criteria, 86.5% of patents are held by large entities. Consistent with Bessen (2008), I find large entities are more likely to renew patents than small entities.

Application Year: Restricting attention to patents renewed at 12 years in 2012 or before (2012 is the last year I have data on pair counts and mainline-set counts) restricts my data to patents granted up through 2000. The average application year of these patents is 1992. I include application year as a control variable and find the probability of renewal is increasing over time.

Application Year > 1995: Patent life was extended from 17 years to 20 years in 1995, and so renewal means the patent stays in force for 5 more years if the application was made before June 8, 1995, or 8 more years if after this date. To account for this change in the value of renewal I include a dummy variable for applications made after 1995. Surprisingly, the coefficient is negative, though we can only reject the null that it is equal to zero when we include fixed effects.