Available online at www.sciencedirect.com
ScienceDirect
Research Policy 36 (2007) 274–287
research policy
www.elsevier.com/locate/respol

The structure of invention

W. Brian Arthur

Santa Fe Institute, 1399 Hyde Park Road Santa Fe, NM 87501, USA

Received 19 December 2005; received in revised form 16 August 2006; accepted 17 November 2006
Available online 9 January 2007

Abstract

This paper explores the process by which radically novel technologies – ones such as radar, the turbojet, or the polymerase chain reaction – come into being. It shows that this process – "invention" – has a certain logical structure common to all cases. Invention is a process of linking some purpose or need with an effect that can be exploited to satisfy it. It may begin with a purpose or need for which existing methods are not satisfactory; this forces the seeking of a new principle (the idea of an effect in action). Or it may begin with a phenomenon or effect itself – usually a freshly discovered one – for which some associated principle of use suggests itself. Either way, translating this base principle into physical reality requires the creation of suitable working parts and supporting technologies. These raise their own challenges or problems, the solution of which may raise further challenges. As a result, invention is a recursive process: it repeats until each challenge or problem (and subproblem, and sub-subproblem) resolves itself into one that can be physically dealt with. It is challenging, usually lengthy, part-conceptual, and part-experimental.

Keywords: Innovation; Technological change; Invention

1. Introduction

Schumpeter famously divided technological change into three phases: invention (the creation of new technologies); innovation (the commercial introduction of new technologies); and diffusion (the spreading of new technologies). Of these invention has been by far the least studied. In this paper I want to take up this lacuna. I will not be concerned with invention as the mere improvement of existing technologies; rather I am interested in it as a process by which radically new technologies come to exist as entities that depart in some deep sense from what went before.

Theories of invention are scarce in the literature but by no means absent. In the 1920s and 1930s a small group of American writers – Ogburn (1922), Usher (1929), Kaempffert (1930), Gilfillan (1935), and others, reacting against the transcendentalist view that the emergence of radical novelty was the work of great men of "genius," stressed that invention was a process. It was a process whereby novel technologies came into being as fresh combinations of existing ones. (Thurston in 1889 and Schumpeter in 1911 had made this point earlier, but neither had enlarged it into a theory of invention.) Usher's version, the most interesting of these, was based partly upon gestalt experiments carried out by Wolfgang Köhler (1926). Köhler had chimpanzees attempt to reach fruit that had been placed out of reach by hand. Under the right conditions, as an act of insight, they picked up and used a stick placed nearby. Generalizing this to the human context, Usher saw at the heart of invention an act of individual insight in which the essential solution to the problem was arrived at. This act had four steps:

> The first step is the perception of a problem, which is conceived of as an incomplete or unsatisfactory pattern. Typically, the problem is an unfulfilled want. Gratification is made effectively possible by some for-

0048-7333/$ – see front matter
doi:10.1016/j.respol.2006.11.005

> tuitous configuration in events or in thought, which present to the individual all the data essential to a solution. This [second step] can be called the *setting of the stage*. . . . [F] or the general process of invention this step is dependent upon pure chance; or upon the mediated contingency of a systematic effort to find the solution by trial and error. . . . The setting of the stage leads directly to the act of insight [third step] by which the essential solution of the problem is found. But this does not bring the process to an end. Newly perceived relations must be thoroughly mastered, and effectively worked into the entire context of which they are a part. The solution must, therefore, be studied critically, understood in its fullness, and learned as a technique of thought and action. This final stage can be described as critical revision.

This as we shall see is not entirely wrong. But by making the core of invention an act that depends upon "some fortuitous configuration in events or in thought," it is frustratingly vague. A more recent contribution by Schmookler (1966) does not improve matters. Again combination is central, and again a series of steps is outlined, but the one at their core is: "The creation or recognition by the originator of the root idea of the invention." Once again this is not wrong, but it begs an account of how such root ideas are arrived at. There is little in these theories by which the invention say of radar, or the polymerase chain reaction, or the turbojet could be compared or individually categorized or better understood.[1]

A different suggestion, one which recurs regularly in the literature, is that novel technologies arise from some process of variation of old technologies and selection of the fittest of these (Basalla, 1988; Constant, 1980). This idea has a certain Darwinian appeal, and it has validity with respect to improvements in technologies. But it does not hold up for what interests us here: radical invention by deliberate human design. Radar certainly did not emerge from the random variation of 1930s radio circuits. A version of variation and selection does apply to technologies such as social institutions or trading contracts that have no deliberate originator. These tend to emerge via a process that has been well studied: One of several practices may gain adherents, partly by random events, which makes it more prevalent. Other members of the community find it advantageous to conform with the more prevalent practice. It may then "emerge" as a social pattern and lock in (Arthur, 1989, 1994; David, 1985).

A significant difficulty that all theories face is that modern research shows that the actual process of invention varies greatly from historical case to historical case, so that universalities appear not to exist. Some novel technologies issue from an individual working alone, others from several groups working with independent ideas. Some derive from huge programmatic investment, others from private shoestring effort. Some emerge from years of trial and are marked by a sequence of intermediate versions that do not quite fulfill the goal, others appear whole cloth as if from nothing. "Attempts thus far to present a general interpretation of all technology change have foundered on the great diversity and complexity of that change," says Constant (1980). As a result, in modern times the idea of "invention" has assumed a status like that of "consciousness" or "mind," something we can speak of but not quite articulate. Textbooks hurry past it without explaining what it is.

About the conditions that foster inventive activity we are much better informed. We know that novel technologies are shaped by social needs (Bijker, 1995); that they respond to economic opportunities, perceived risk, and factor price changes (David, 1975; Dosi, 1988; Freeman, 1990; Rosenberg, 1982); that they cumulate with the accretion of cultural and scientific knowledge (Mokyr, 2002); and that they can be catalyzed by the exchange of information within networks of colleagues (Aitken, 1985; Lane and Maxfield, 1997).

In this paper I want to look in detail at the process by which radically novel technologies originate. I will build on the idea that combination is central. My purpose is to show that invention has a certain logic or structure to it, and that indeed this logic explains why and how the process varies. I will argue that invention is a process of linking some purpose or need with an effect that can be exploited to satisfy it. It proceeds from a need for which existing methods are not satisfactory, which forces the seeking of a new principle (the idea of an effect in action); or from a phenomenon or effect itself – usually a freshly discovered one – for which some associated principle of use suggests itself. Either way, translating this principle into physical reality requires the creation – and combination – of suitable working parts and supporting technologies. These raise their own challenges or problems, the solution of which may require inventions of their own. The process, I will argue, is one of problem solving – recursive problem solving.

My argument will depend upon carefully defining what a technology is and how it is structured. I will do this in Section 2. Section 3 will explore what it means

[1] For commentary on the 1920s theories see McGee (1995), and Ruttan (1961). For further studies on invention see Jewkes et al. (1969), and Kaiser (1999).

to be a radically "novel" technology. Sections 4 and 5 will describe in detail the process of invention; and Section 6 will comment upon this. Section 7 will examine the wider context of invention, namely the cumulation of technologies and knowledge that precede any novel combination. Some caveats are in order. Throughout I will be proceeding at a micro-process level, and hence will avoid the usual economic discussions of demand-pull versus technology-push, factor-saving biases, or induced innovation. Also, while I talk here about the creation of radically novel technologies, I recognize that the step-by-step improvement of existing technologies is economically just as important. And, like the 1920s writers, I will use the word "theory" to describe systematic accounts of the invention process (including this one). One meaning of *theory* (Webster) is "a systematic statement of the principles involved," which is exactly what we are seeking.

2. How technologies are structured

I will define *a technology* in this paper quite simply as a *means* to fulfill a human purpose. The purpose may be explicit: to power an aircraft say, or to sequence a DNA sample; or it may be hazy, multiple, and changing: a computer has no single, explicit purpose. But whether its purpose is well defined or not, a technology is a means to carrying out a purpose.[2] A power station supplies electricity. The Haber process produces ammonia. As a means to fulfill a purpose, a technology may be a method or process or device: a particular speech recognition algorithm, say, or a filtration process in chemical engineering, or a type of diesel engine.

Technologies are put together or *combined* from component parts or assemblies. A hydroelectric power generator combines several main components: a reservoir, an intake system with control gates, an intake sluice or penstock, turbines, electricity generators driven by the turbines, transformers to convert the power output to higher voltage, and an outflow system. Such assemblies, or subsystems, or subtechnologies (or stages, in the case of process technologies) are groups of components that are largely self-contained and largely partitioned off from other assemblies. Of course some very simple technologies – a rivet, for example – may have only one component part.

[2] The word technology has two other legitimate meanings: a body of practices and components, such as electronics or optical data transmission; and "the totality of the means employed by a people to provide itself with the objects of material culture (Webster)."

A technology always proceeds from some central idea or concept – "the method of the thing." I will call this the *base concept* or *base principle* of the technology. Thus the base principle of a clock is to count the beats of some stable frequency. The principle of radar – the essential idea that allows it to work – is to send out high frequency radio waves and detect distant objects by analyzing the reflections of these signals from the objects' surfaces. The principle need not be simple. The base principle of xerography is to electrostatically charge a surface (a copier drum usually) and project light from an image onto this surface so as to discharge the areas illuminated; opposite-charged toner particles will then stick to the still-charged areas (so that the dark parts of the image remain on the drum) and can be rolled onto paper and heat fused there.

A principle is merely an idea, a concept. It must proceed from something – something usable or exploitable. It proceeds always from an effect, some phenomenon (or set of phenomena) that it exploits. I say "always" because a technology that exploited nothing could achieve nothing. Thus a clock exploits the physical phenomenon that certain objects – pendulums or quartz crystals or ammonia molecules – oscillate at a fixed frequency (Landes, 1983, 1989). An aircraft jet engine (or gas turbine powerplant, to give it its proper name) exploits the phenomenon that a mass expelled backward produces an equal and opposite forward reaction. The phenomenon exploited need not be physical. Any reliable or repeatable effect from nature, or logic, or behavior, or organization may be harnessed for use.[3] Of course, sometimes more than one effect or phenomenon is central to a technology. A laser printer embodies the base concept that a computer-controlled laser can "paint" high-resolution images onto a copier drum, and to do this it harnesses two equally important phenomena: that a laser can produce very intense highly-focusable light, and that a charged surface can attract oppositely charged particles. To simplify the exposition that follows I will usually speak as if a technology exploited a single central effect (and a physical one).

A technology, then, is built around the reliable exploitation of some effect, as envisaged through some principle of use. In practice this means that a technology consists of a main assembly – an overall backbone of the

[3] For example, the heterodyne principle in radio stems from a mathematical effect or truism: If two sine waves of different frequencies are combined, the resultant wave "beats" (oscillates) at new frequencies equal to the difference and sum of the original ones. In a typical application, this effect is used to shift the frequency of a received signal by "mixing" it with an internally generated wave.

device or method that executes its base principle – plus other assemblies hung off this to support its working, regulate its function, and feed it with energy. These in turn require their own sub-components and sub-assemblies: controlling mechanisms, monitoring devices, input and output interfaces. Thus a jet engine has a main assembly that consists of an air intake system, a compressor system (to compress the inducted air for combustion), a combustion system (to provide high-energy gas flow for the turbine), a turbine system (to drive the compressor and provide reactive thrust), and an exhaust section. Each of these in turn is controlled, supplied, and monitored by other subsystems: the compressor system requires a variable vane actuating system (to set the vane angles appropriately to airflow velocity), and an anti-stall bleed system to control pressure surges (the tendency of the compressed air to blow backwards); the turbine system requires a blade cooling system, and a complicated set of shrouds and seals to prevent high-pressure gas leakage (Otis, 1997).

This picture of a technology consisting of components that consist of further components gives us one other property I would like to define. Each component system or assembly of a technology itself has a purpose, an assignment to carry out. If not, it would not be there. By my earlier definition each assembly is therefore itself a means – a technology. And each assembly has its own subassemblies or components. Each of *these* in turn has an assignment to carry out. Each also is a means – a technology. This pattern, that a technology consists of building blocks that are technologies, that consist of further building blocks that are technologies, repeats down to the fundamental level of individual components. I will call it *recursiveness*. Practically speaking it means that a technology is organized in a loose hierarchy of groupings or combinations of parts and subparts that themselves are technologies. This hierarchy can be as many as five or six layers deep. Of course the hierarchy is not perfectly tree-like, with a main trunk and branches leading off from branches and branches from them in turn, because subsystems can interact and crosslink at different levels. Recursiveness implies that properties that apply at one hierarchical level of technology apply at another level, and therefore that each component at each level also has a base concept and exploits some effect. And, importantly for us, it will imply that invention is not a matter of solving problems at the level of an overall technology. It will be a matter of going back and forth among levels, taking care of problems at one level by finding solutions at another level.

To summarize, a technology is a means to fulfill a purpose, and it does this by exploiting some effect. It consists of a central assembly – the overall backbone of the device or method that executes its base concept (and exploits one or more base effects) – plus other assemblies hung off this to make this workable and regulate its function. These components or assemblies function together in a *working architecture*.[4] To understand a technology means to understand its principle and how this translates into components that share a working architecture.

In the argument that follows it is important that the reader keep a clear distinction between phenomenon and principle. That air pressure falls with altitude is a physical phenomenon; the idea of using this effect to measure altitude constitutes a principle. That radio waves are reflected from metal objects is a phenomenon; the idea of using this to detect metal objects at a distance (in radar) constitutes a principle. A phenomenon is simply a natural effect, and as such it exists independently of humans and of technology; it has no "use" attached to it. A principle by contrast (as I will use the word) is the *idea of use of a phenomenon for some purpose*, and it exists very much in the world of humans and of use. This idea of use may be specific to a technology: the principle of xerography. Or it may be generic, attached to a phenomenon and therefore available across many technologies: the principle of reactive propulsion.

3. What exactly is invention?

We are now ready to talk about invention. What exactly is it? What, in our context, allows one new technology to qualify as radically novel, and relegates another to be a mere improvement on or variation of some standard design?

A technology, as we have just seen, possesses a purpose, a combination of components, an architecture, and embodies a base principle that exploits some base phenomenon. We could therefore define an invention to correspond to a significant change in any one of these. A little thought shows that the first three are not usually what we think of when we recognize a change in technology fundamental enough to qualify as radically novel. Personal computers change their purposes all the time, their components some of the time, and their architecture occasionally, but they remain personal computers. Such changes in purpose or components or architecture might

[4] A technology need not be fixed in configuration. Most technologies in fact are adaptable in architecture, constantly changing in configuration and purpose as differing needs require. But basic to my argument is the assertion that all technologies share the common organizational structure I have outlined above. For a fuller account see my forthcoming book (Arthur, 2007).

imply a *modified* technology, but not a radically novel technology.

A change in a base *principle* by which the purpose is achieved, however, is a better candidate. A principle, remember, is the idea of use of some central effect. When we designate a novel technology as an "invention," we find always a purpose carried out by a new or different base principle. Consider: In the 1930s approaching aircraft could be detected over the horizon by listening for acoustic emissions. Radar was based on a different principle: picking up the faint echoes that aircraft reflected from radio pulses. In the 1970s computer printing was carried out by the line-printer with its limited set of fixed characters. The laser printer was based on a different principle: using a laser to "paint" images – any image a computer could produce – onto a copier drum. In the 1930s aircraft could be powered by a piston-and-propeller system. The turbojet was based on a different principle: using a constant airflow powered by a gas turbine to provide reactive thrust. In each case the new technology exploited a different principle, new to the purpose in hand.

A change in principle, then, fits with our intuition of what constitutes a novel technology. I will therefore define a new (radically novel) technology as one that achieves a purpose by using a new or different base principle than used before. This has the right feel to it. We can say that Watt's steam engine is an improvement of Newcomen's, not an invention. It provides for a new component – a separate condenser – but not a new base principle. (Watt's case proves that sometimes improvements can be more significant economically than pure inventions.) And our definition properly allows for gray areas; often the newness of a principle to a particular purpose is debatable.

Thus far we merely have a criterion that separates invention from mere modification, not an explanation. It is tempting to assume that novel technologies arise from envisaging a different principle applied to some purpose. This is certainly part of the story. But invention is more complicated and more varied than that, and we need to look at it more closely.

Recall the three properties of technologies I outlined earlier: (1) A technology fulfills some expressed purpose – some need – personally or socially perceived. (2) A technology is built always around the reliable exploitation of some base phenomenon as envisaged through some principle of use. (3) A technology requires other sub-principles (and therefore sub-components) for its practical working. It consists of components that are themselves technologies (and that in turn consist of further technologies), the whole arranged in a recursive hierarchy. By the first two properties, a radically novel technology must link some purpose or need with an effect that can be exploited to satisfy it (using by our criterion a novel principle). This linkage, as we will see, is a process, not an event. And it is this usually difficult process that constitutes invention. The process may begin with a purpose or need for which existing methods are not satisfactory; this forces the seeking of a new principle (the idea of an effect in action). Or it may begin with a phenomenon or effect itself – usually a freshly discovered one – for which some associated principle of use suggests itself. Either way, translating this base principle into physical reality requires the creation of suitable working parts and supporting technologies (property 3). These raise their own challenges – indeed some may require inventions of their own. As a result, invention is primarily a process of recursive problem solving. It entails matching a need to a principle (or effect envisaged in use) and solving the hierarchy of problems and subproblems that this creates. Indeed, often solving these subproblems constitutes the bulk of the work.

Within this overall structure there still exists considerable scope for variation. Sometimes the process is accomplished by a modest effort; in other cases it requires exertions on a national scale. Sometimes it requires deep theoretical understanding of the phenomenon used; at other times the challenges are more practical and experimental. The possible variations are many.

In the next two sections we will look at this process of invention in some detail. For convenience I will present it mainly as experienced by a single individual (or group) who takes it from beginning to end. But we should keep in mind that very often many efforts are under way simultaneously, with some originators ignorant of others and some borrowing from others; and that the process may be split among practitioners, some taking it part way and others building on these earlier efforts. We should also bear in mind that behind the people-driven process I describe lies a deep set of supporting causal factors: of antecedent ideas and understandings of phenomena, and of previously developed components and principles. I will briefly talk about these later.

4. The base conception

4.1. *The process initiated from a need*

Let me begin by first looking at the pattern where invention starts from a particular need. The need in question may arise from an economic opportunity, the recognition of a potentially lucrative market perhaps; or from a change in economic circumstances; or from a social challenge; or from a military one. After the First World War, military aircraft improved rapidly in range and speed, and by the early 1930s Britain became acutely aware of its vulnerability to attack from the air. The menace became the subject of political and public debate. "That there is at present no means of preventing enemy bombers from depositing their loads of explosives, incendiary materials, gases, or bacteria upon their objectives I believe to be true," wrote Frederick Lindemann to The Times in 1934. The British Air Ministry took notice and considered different principles to respond to the problem (Buderi, 1996). Among these, and by no means the most promising at the beginning, was the idea of detecting aircraft by reflected radio waves – what was subsequently called radar.

Often the need arises not from an outside stimulus, but from within technology itself. In the 1920s aircraft designers realized they could achieve more speed in the thinner air at high altitudes. But at these altitudes reciprocating engines, even when supercharged, had trouble drawing sufficient oxygen, and propellers had less "bite." Needed was a different principle than the piston-propeller one.

Typically the need sits for some time with at least some practitioners aware of it, but with none seeing an evident solution. If there were one, standard technology would suffice. The question is therefore by definition challenging. Those that do take the challenge (I will call them originators, to avoid the lone eccentric connotation of "inventors") may encounter the situation as a need to be fulfilled or a limitation to be overcome; but they quickly reduce it to a set of desiderata – a problem to be solved. Both Frank Whittle and Hans von Ohain were aware of the limitations of the old piston-and-propeller principle and of the need for a different one; but they re-expressed these as a technical problem – a set of requirements to be met. Whittle sought a power unit that was light and efficient, could compensate for the thin air at high altitudes, and could if possible dispense with the propeller. And von Ohain sought a "steady aerothermodynamic flow process" noting that "the air ducted into such a system could be decelerated prior to reaching any Mach-number-sensitive engine component" (Constant, 1980). The need or limitation becomes a well-specified problem.

The problem now comes forward as it were, looking to meet an appropriate solution. The mind (for the moment I'll treat the originator as a singular mind, but more usually several minds are at work) becomes fixed on the problem. It scans possibilities that might with further development satisfy the desiderata. This search is conceptual, wide, and often obsessive.

What is being sought at this stage is not a full design, not a full architecture along with the components that it will fulfill it. What is being sought is a base concept – a principle – the idea of some effect (or combination of effects) in action that will fulfill the requirements of the problem, along with some conception of the means needed to achieve this. A conception of these supporting means is necessary because each candidate principle when considered seriously brings up its own particular difficulties and these pose subproblems. Such difficulties narrow and redefine what needs to be solved, as the mind realizes that if a certain principle is to be achieved, a certain component piece is necessary; or if a component piece can be achieved, the larger solution will follow. Thus the process goes back and forth between levels, testing the feasibility of principles at one level and attempting to deal with the problems these raise at a different level.

The process here resembles the way a route up an unscaled mountain might be planned. To reach the summit is to solve the problem. And to envision a base principle is to posit a promising overall route or major parts of a route, with a given starting point. On the mountain are patches of obstacles: ice falls, awkward traverses, head walls, stretches subject to avalanches and falling rock. The solution can be plotted from the top – the requirements of overall problem – down. Or from the base – the requirements of the overall principle – up. Each new principle or overall plan of climb meets its own difficult stretches that must be got past. Here recursiveness comes into play, because each obstacle stretch becomes its own sub-problem and requires its own solution (or sub-principle or sub-technology, in our case). An overall solution is not achieved until some starting point at the base is connected in a reachable way with the summit. Of course, certain stretches of the mountain may have been climbed before – in our context certain sub-technologies may be available and the solution will be biased toward using these. So the process may be more like stitching together known parts than pioneering a complete route from scratch. Each piece (a problem) must be met with a route to deal with it (a principle). And each obstacle on the overall route must be met with its sub-route (sub-principle) to deal with that. The process is in part recursive and the whole becomes a concatenation of parts, a combination of stretches. It forms a plan of advancement, or in our case the envisioning of a technology.

With technology the candidate routes are not visible, and must be sought by other means. Where do these candidate routes – these principles – arise from? Sometimes, as with the birth of radar, a fresh phenomenon is conveniently at hand to supply a base principle. But more usually principles are borrowed, appropriated from other purposes or devices that use them. Whittle, in 1929, mulled through the possibilities of rocket propulsion, reaction propulsion using a rotating nozzle, turbine propulsion using a propeller (a turboprop), and a ducted fan blower (a reaction jet) powered by a piston engine – all the while pondering the subproblems these would raise (Constant, 1980; Whittle, 1953). Each of these possibilities was borrowed from technologies used for other purposes. Sometimes a new overall principle is suggested by combining two or more borrowings. Randall and Boot hit on the principle of the cavity magnetron – a cylindrical electron tube used to generate microwaves for radar purposes using a magnetic field to control the electron flow – by combining the positive aspects of the magnetron (its high power output) and of the klystron tube (the idea of resonant cavities). Sometimes a principle is recalled from the past, or picked up from the remark of a colleague, or suggested by theory. Indeed Randall's recent encounter with an English translation of Hertz's *Electric Waves* had suggested to him the notion of a cylindrical resonant cavity – basically a three-dimensional version of the wire loop resonator Hertz analyzed in his book (Burns, 1988). However principles are arrived at, they are never invented from nothing. They are appropriated from or suggested by that which already exists, be it other devices or methods or theory or phenomena. This process of mental appropriation and half-conscious suggestion lies at the creative heart of invention.

Occasionally the sought-for solution, the conceptual combination that eventually proves successful, is arrived at by systematic investigation of the possibilities. "I therefore started to examine systematically all possible alternative methods" says Francis Aston of his explorations that would lead to the mass spectrograph (Aston, 1922). But more often the mulling of principles and the considering of means to resolve the technical obstacles they present goes on unsystematically. It persists for some time, with several false starts, or with possibilities stymied by some obstruction. Then the overall problem sits unresolved. It may be pushed to the back of the mind, temporarily left to itself.

The solution may arrive abruptly. "The key revelation came in a rush," says Charles Townes (1999), of his insight into what would become the maser. "Suddenly I knew how to do it," says Kary Mullis (1999), of his concept that would become the polymerase chain reaction for amplifying DNA samples. And Whittle (1953) says:

> While I was at Whittering, it suddenly occurred to me to substitute a turbine for the piston engine [to drive the compressor]. This change meant that the compressor would have to have a much higher pressure ratio than the one I had visualized for the piston-engined scheme. In short, I was back to the gas turbine, but this time of a type that produced a propelling jet instead of driving a propeller. Once the idea had taken shape, it seemed rather odd that I had taken so long to arrive at a concept which had become very obvious and of extraordinary simplicity.

The insight comes as an overall principle with a workable combination of sub-principles, or as a sub-principle that clears the way for the main principle to be used. It comes a moment of connection, always a connection, because it connects a problem with a principle – an effect in use – that can handle it. Strangely, for people who report such breakthroughs, the insight arrives whole, as if the subconscious had already put the parts together. And it arrives with a "knowing" that the solution is right – a feeling of its appropriateness, its elegance, its extraordinary simplicity. The insight comes to an individual person, not to a team, for it wells always from an individual subconscious. And it arrives not in the midst of activities or in frenzied thought, but in moments of stillness.

What has been arrived at is a concept by which to work. This arrival is not the end of the process; it is merely a marker along the way. The concept must still be translated into a working prototype of a technology before the process is finished.

4.2. *The process initiated from a phenomenon*

So far the base principle is arrived at from a need. When the process begins from the other end of the linkage, from a phenomenon (usually the discovery of a novel one), the base principle is arrived at differently. It is *suggested* by the phenomenon rather than *sought* from a need. Roentgen's accidental discovery of X-rays in 1895 almost immediately suggested the principle of using these to illuminate bones and tissues inside the human body. Indeed the article Roentgen circulated within two months of his investigations contained striking pictures of his wife's skeletal hand, so the principle was both public and obvious (Kevles, 1997).

It would seem that things should be simpler here. If the principle is suggested by the phenomenon, the difficult task of searching for a base principle is eliminated. But in most cases the suggestion is far from automatic. It is one thing to notice a phenomenon and a different thing to derive a clear principle from it and pursue this with the intention of creating a technology. Fleming in 1928 noticed the effect that a substance within a mold (spores of *Penicillium notatum*) inhibited the growth of a culture of staphylococci bacteria. But others had noted the phenomenon before him – John Tyndall in 1876 and André Gratia in the 1920s, for example (Lax, 2005; Clark, 1985). Unlike them, Fleming clearly articulated a principle of use and undertook systematic experiments to construct a therapeutic means from it.

Even when a principle *is* clearly articulated, a phenomenon may sit for several years before it is translated into a working technology. A pressing need may be missing, or formidable obstacles may lie in the way of capturing the phenomenon for use. Translating the *Penicillium* effect into a working technology required that the active substance in the mold be isolated, purified, and stabilized; that its chemical structure be characterized; that its curative properties be demonstrated; and that methods for production be developed (Hare, 1970; Chain, 1971; Williams, 1984; Lax, 2005). All this called for more specialized types of expertise than Fleming possessed, and it constituted a new phase in the invention process. It fell to a team of biochemists led by Florey and Chain at Oxford's Dunn School of Pathology to carry out this phase. The gap between Fleming's initial observation and the emergence of usable penicillin was 13 years.

Translating a principle into a workable technology is indeed a new phase, whether the principle has been arrived at by seeing the possibilities of a phenomenon or by pondering the requirements of some need. The process must now be taken from mental concept to physical embodiment, and this gives it a more physical character. Solutions that were conceptual must be produced in physical form, and subproblems that were partially bypassed must be dealt with directly. All this requires considerable effort, and as in the Fleming case, is often accomplished by a different group of people.

5. From principle to working technology

The new phase normally will have been already partially under way. Some components of the device or method may have been constructed in experiments, and physical trials of the base concept in action may have been attempted. But even with such early results in hand, challenges still arise. Envisaged subtechnology solutions may not work, or may press upon performance limits. Whittle faced combustion difficulties in his early tests. The combustion chambers tended to overheat and distort, soot formation "coked up" the vaporizer tubes, and the distribution of temperatures at the combustion chamber outlets was uneven (Whittle, 1953; St. Peter, 2002). His designs required compression ratios that lay beyond current standards. Such subproblems can normally be handled by stretching standard engineering – they were in Whittle's case – but others may themselves call for radical solution. Indeed, the most important contribution of the British radar effort lay not in envisioning of the principle of radar. That had been seen by many in the scientific community before.[5] It lay in solving a critical sub-problem, that of finding a means for producing high-powered microwave signals, by originating a component technology – the cavity magnetron.

Not infrequently resolving subproblems requires efforts that dwarf those required for arriving at the base principle. Gary Starkweather had seen the central concept of the laser printer – the idea of using a laser to paint an image on a Xerox drum – early on. Indeed the idea was in the air: George White at Electro-Optical Systems, for example, had experimented with the principle (Hiltzik, 1999). But to make the concept a working reality, Starkweather faced several difficulties. Commercial considerations required that a page of written text be scanned onto a copier drum in at most a few seconds. If this was to be achieved with high resolution, the laser beam would need to be capable of being modulated (switched on and off to mark black or white dots on the drum) at the rate of 50 million times per second (Starkweather, 1997; Hiltzik, 1999). Further, the photoconductor coating of the drum was thought to suffer fatigue (become less sensitive) over time if exposed to intense laser light. And any laser and lens module would be too heavy – have too much inertia – to be mechanically moved back and forth thousands of times per second as required to scan lines onto the drum. Each of these problems needed to be resolved before a working technology could be accomplished. Starkweather solved his modulation problem by developing a very fast shuttering device using a polarizing filter driven by a piezoelectric cell. He resolved the fatigue issue (it turned out to be false) by an extensive series of tests. He solved the inertia problem by keeping the laser module stationary and moving only the beam using a rotating multifaceted mirror. Each mirror facet could scan a thin line across the drum as the mirror revolved. But this solution brought it own sub-subproblem. Adjacent facets of the mirror, Starkweather calculated, would need to be vertically aligned to a tolerance of 6 arc-seconds, else adjacent scan lines

[5] As early as 1904, a German engineer, Christian Hülsmeyer, had taken out patents on a device for preventing collisions at sea, using radio waves, and by the 1930s several practitioners, Marconi among them, had experimented with primitive radio detection devices (Süsskind, 1988, 1994).

would not be properly offset and the image would be distorted (Starkweather, 1980, 1997). But the costs of machining to such precise tolerances were prohibitive. A carefully designed cylindrical lens – Starkweather's main expertise was optics – took care that adjacent lines fell close even if the mirror facets were slightly misaligned. Each such subproblem required a non-standard solution – a mini-invention of its own – with attendant trials of alternative methods, failures, and long sequences of experimentation.

It is by no means unusual that invention consists largely of solving subproblems. Indeed, often the base principle has been established some years in the past but sits stymied by technical obstacles. The most visible part of the process then consists of solving these subproblems. This was the case with the Manhattan atomic bomb project.[6] The base concept was well known by the late 1930s: many groups of physicists were aware that a self-sustained nuclear chain reaction could be used as a powerful source of energy. Indeed, Leo Szilard had conceived of the chain reaction concept as early as 1933 and had bruited it widely within the physics community (Rhodes, 1986). But this principle remained nothing more than a scientific vision until the pressures of war called it into material being. (Here we can say that both need and phenomenon initiated the technology.) Between principle and purpose lay formidable obstacles: technical subproblems that required inventions of their own.

Chief among these was a means to separate the fissionable isotope U235 from the chemically similar U238 isotope. Various methods were proposed: the fissionable material could be separated by centrifuge, by electromagnetic separation, by gaseous barrier diffusion, or by liquid thermal diffusion (Rhodes, 1986; Badash, 1998). Each method had its own proponents, and its skeptics. And each had its own principle. Thermal diffusion was based on the principle that lighter isotopes tend to migrate toward a hotter region and heavier isotopes to a cooler one; gaseous barrier diffusion on the principle that lighter molecules tend to diffuse through a porous barrier faster than heavier ones. No method at the start was much more than a proposal for reaching a higher proportion of U235: even under ideal conditions a gaseous diffusion unit could enrich uranium by a factor of only 1.0043, so its process required a cascade of thousands of interconnected units. As each method moved into pilot program stage it encountered its own technical obstacles. Uranium in gaseous hexafluoride form proved to be highly corrosive; it attacked seals made of organic material in pumps or pipe connections. This required radical solution, which came from developing a plastic seal made of a new material, Teflon. In the end these and other lower level obstructions were cleared, and after major efforts a combination of separation methods delivered the product.

When invention consists mostly in finding working solutions to challenging subproblems, as in this case, the process has more the character of development. Precisely focused effort is more usually required than conceptual breakthrough, and so here we rarely see moments of epiphany. Solutions are proposed – and fail; parts do not work; redesigns are necessary; and endless tests must be made. The process becomes a progressive advance across a broad front as knowledge is gained and subtechnology challenges are successively resolved, pressing always toward a version that works properly.

The first pilot device to do this is always an achievement. Even if its initial showings are feeble, the moment nonetheless is precious. The thing works and a milestone has been passed, to the jubilation of those present. The initial demonstration may indeed be weak, but with further efforts and ad-hoc fixes – and subsequent versions with better components – a robust working version emerges, and the new base principle comes into a semi-reliable state of being. It has taken physical form. All this takes time – time that tries the patience of backers and supervisors. And time in which most necessary human ingredient is will, the will to bring the principle to life as a working entity. Now the new device or method becomes a candidate for development, and commercial use. It may, if it is fortunate, enter the economy as an innovation. Invention, as a process, is now complete.

6. Discussion

Let me summarize at this point. Invention is at bottom a linking of some purpose or need with an effect that can be exploited to satisfy it. It falls into two overlapping phases: the search for a principle (or the suggestion of one from a phenomenon or effect); and the translation of this into physical reality. Both phases bring up challenges, the solution of which may raise further challenges. The process is therefore recursive: it repeats until

[6] In this instance, the subproblems were so challenging their solution has come to be thought of as the actual invention. This is also the case with powered flight. By 1900, the two base principles of powered flight (propulsion via internal combustion and lift from fixed-wing airfoils) were known and accepted. The Wright brothers solved the four main subproblems needed to achieve this: providing mechanisms for control and stability; finding wing sections with good lift; developing a lightweight, efficient powerplant; and developing a high efficiency propeller.

each challenge or problem (and subproblem and sub-subproblem) resolves itself into one that can be dealt with using existing elements.

What exactly are these elements – the building blocks – that originators use? Ostensibly they are already existing technologies in the form of components, assemblies and methods. But conceptually, in the originator's mind, they will more likely be thought of as *functionalities*: generic actions or operations that lie at hand. (By a generic operation here I mean one usable in a variety of contexts.) Thus electronics designers know that they can translate a high frequency into a lower one by mixing it with a fixed frequency; that they can smooth a signal by using a capacitor in parallel; that they can get rid of the DC component of a signal by a capacitor in series; and that they can make use of a hundred other reliable effects. Such functionalities are the currency of invention.

Thus Lawrence in inventing the cyclotron (which accelerates charged particles to high energy) does *not* immediately think in terms of combining an electromagnet with an oscillating electric field between two D-shaped containers. He *knows*, as any physics student does at the time, that electric fields can accelerate charged particles (a functionality). And he *knows* you can use a magnetic field to cause charged particles (ions) to travel in circular paths (another functionality). He has been vaguely searching for a means to accelerate ions, but the various proposals in currency are obstructed by the problem of achieving the extremely high voltages necessary to provide the accelerating field. He notices in a German journal (*Archiv für Elektrotechnik*) Wideroë's suggestion to send ions through a series of tubes laid end to end using relatively low AC voltage applied across the gaps between the tubes to accelerate them piecemeal (Lawrence, 1951; Wilson, 1998; Wideröe, 1928). (These are arranged so that the ions arrive at the gaps just as the AC voltage peaks.) But Lawrence calculates that to achieve the energies that he wants, the series of tubes would be impracticably long. He realizes – and this is the inception of his cyclotron principle – that the particles do not need to travel down a series of tubes. Instead he can use two tubes over and over if they are bent to form two halves of a circle separated by gaps, with the ions forced by a suitable magnetic field to circle repeatedly within them. Wideroë's well-timed voltages can be applied across the gaps between the two tubes to accelerate the ions each time they cross from one tube to the other. As the ions circle they will pick up velocity and gradually spiral outward (the tubes can be widened to form two semicircular containers to accommodate this), eventually to be led off for high-energy use (Lawrence, 1951). Originators think in terms of achievable actions and deliverable effects – what I am calling functionalities – and they combine these in solving problems.

Functionalities of course are also the currency of standard technological design. But what differentiates invention is that the overall problem has not been satisfactorily solved before, that the challenges may run several recursive levels deep, that the solutions of these may be far from standard, that novel phenomena and unusual effects may have to be used, and that the overall principle is new to the purpose in question. All these add to difficulty, but they do not make invention qualitatively different from design.

By this reasoning, what is common to originators is not "genius" or special powers. Rather it is the possession of a very large quiver of functionalities. Originators are steeped in the practice and theory of the principles or phenomena they will use. Starkweather chose holography as his doctoral topic, and therefore was adept with the theory and practice of lasers; his master's degree was in classical optics. Whittle's father was a machinist and inventor, and Whittle was familiar with turbines from an early age. Originators, however, do not merely master functionalities and use them once and finally in their great creation. What always precedes invention is a lengthy period of accumulating functionalities and of experimenting with them on small problems. Often in this period we can see hints of what they will use. Five years before his revelation, Townes had argued in a memo that "[m]icrowave radio has now been extended to such short wavelengths that it overlaps a region rich in molecular resonances, where quantum mechanical theory and spectroscopic techniques can provide aids to radio engineering" (Buderi, 1996). We can see this cumulation of functional expertise in what originators take for granted. Mullis (1999) remarks on the simplicity of his polymerase chain reaction scheme (which reproduces a large number of DNA strands from the very few in a given sample). "It was too easy. . . . Every step involved had been done already." But Mullis's "easy" solution was to "amplify DNA by the repeated reciprocal extension of two primers hybridized to the separate strands of a particular DNA sequence," something easy only to a practitioner with considerable experience of functionalities in working with DNA.

For convenience I have described the process largely as if one individual originator or group was at work, bringing forth something entirely new. But in most cases of invention a group or team is at work, especially in the phase of translating the concept into a working technology. And in most cases we can find some vague prior articulation or prior embodiment of the principle, perhaps not well grasped, but prior just the same.

Further, almost as often we find a series of prototypic versions by different workers who borrow from each other, with the device or method improving gradually in effectiveness from crude beginnings as improved sub-technologies are found. The computer is an example. Many "inventions" are in reality improvements on earlier embodiments of a known idea. Randal and Boot are credited with the invention of the cavity magnetron, but in actuality a decade's worth of experimentation and theorizing on split-anode magnetrons had preceded their device. Indeed Hans Hollman in Germany had been granted a US patent on a cavity magnetron two years before Randall and Boot's work (Thumm, 2001; Süsskind, 1994; and Callick, 1990). This fact, that the principle has occurred to several groups and that different embodiments – different working versions with different degrees of effectiveness – exist, thwarts efforts to assign credit for "being first" or for "invention" to a single person or group. If credit must be assigned it should go to the person or team that first had a clear vision of the principle, saw its potential, fought for its acceptance, and brought it fully into satisfactory use, and often there are several of these.

I have said little about human interaction and informal networks of communication. At every step these greatly enhance the process I have described. They steep the originator in the lore that has built up around the problem and around previous efforts. They provide suggestions of useful techniques and of principles at work in other domains. They help the originator see the problem differently. Lane and Maxfield (1997) talk about *generative relationships* that "can induce changes in the way the participants see their world and act in it and even give rise to new entities, like agents, artifacts, even institutions." Human interaction also provides needed criticism to burst fanciful bubbles, and it provides equipment and know-how to bring the concept to physical reality. Of course, too much interaction can be harmful. This is because a certain degree of obsession is required and cannot be generated in a diffuse group, and because bringing a radically different principle to life requires some isolation from standard thinking.

7. Harnessed phenomena and their consequences

I have described invention as a micro-process; but it is one that occurs in a context. A novel technology may seem to materialize out of nothing, but it emerges always from a cumulation of previous components and functionalities already in place. Thus the maser emerged from a collection of functionalities in place by 1950: the separation of ions via fields, the use of resonating chambers, the use of sensitive high-frequency receivers and detectors, the use of techniques from waveguide spectroscopy; and knowledge of molecular properties and resonance. In fact, supporting any novel device or method is a pyramid of causality that leads to it: of other technologies that used the principle in question; of antecedent technologies that contributed to the solution; of supporting principles and components that made the new technology possible; of phenomena once novel that made these in turn possible; of instruments and techniques and manufacturing processes used in the new technology; of previous craft and understanding; of the grammars of the phenomena used and of the principles employed; of the matrix of specific institutions and universities and transfers of experience that lead to all these; of the interactions among people at all these levels described. This wider perspective, called the combination/accumulation view (McGee, 1995) does not negate what I said earlier; the historical causality is complementary to the micro-process of invention. Nor does it imply that the progression of new technologies is predetermined. Invention is subject to the vagaries and timings of the discovery of new phenomena, of the appearance of new needs, and of the individuals who respond to these. If we think of new technologies in this way, as combinations of existing technologies possibly going on to become building blocks for future descendant technologies, we can see a web of technologies building out over time from primitive ones.[7]

This wider context of invention is by no means amorphous; it has its own structure. All inventions are harnessings of phenomena to a purpose; and phenomena present themselves in natural clusters bringing a train of technologies in their wake. As electricity comes to be understood between 1800 and 1875, a constellation of phenomena presents itself: capacitance, induction, deflection of charges by electric and magnetic fields, glow discharge, and electromagnetic radiation. These bring a train of technologies that includes capacitors and inductors, transformers, telegraphy, the electrical generator and motor, the telephone, wireless telegraphy, the cathode ray tube, the vacuum tube, and in due course modern radio, television, radar, electron microscopy and computers. Similarly, as quantum understandings grow in the twentieth century (Kragh, 2002), the laser, transistor, integrated circuit, magnetic resonance imaging, high-temperature superconductor, carbon nanotube, and other technologies based on these emerge. As a field of phenomena is understood and worked with

[7] See Arthur and Polak (2006) for a model of this buildout.

(a recent example is molecular biology) technologies based upon these follow. The result is a clustering of related devices and methods that form technological revolutions.[8] Recently Mokyr (2002) has pointed out that technologies issue forth as human knowledge is gained. This is certainly true. But I would express this idea differently: technologies issue forth as knowledge of phenomena and their theory is gained, so that novel technologies emerge both from the cumulation of existing building-block technologies and from understandings of the phenomena that surround these.

In fact, knowledge is not quite the right word. Radically novel technologies arise more from a context of knowings: they arise from practice in working with – and knowing in a deep way – certain components and functionalities and certain newly uncovered effects. Such practice is really a form of craft. It consists not just in knowing functionalities and how to combine them. It consists in knowings of what is likely *not* to work, what methods to use, whom to talk to, what theories to look to, and above all of how to manipulate phenomena that may be freshly discovered and poorly understood. (See Brown and Duguid, 2000; and cf. Polanyi, 1967). The Cavendish Laboratory at Cambridge was the locus of inventions in atomic physics in the first three decades of the twentieth century. It built these upon a treasury of knowings to do with atomic phenomena. "Whatever was known in this field – techniques, equipment, mathematical tools, even theory – it was known by someone there," says Cathcart (2004), "and more than that it was discussed, challenged and tested at colloquia and other gatherings. To any problem or difficulty in atomic physics there would surely be an answer somewhere in the [Cavendish]."

This observation has policy implications. Radically novel technologies do not arise from mere access to building blocks, nor from mere investment in laboratories, nor even from mere knowledge of scientific phenomena. These of course are important (Murmann, 2003); but they are not sufficient. National leadership in the creation of advanced technologies issues from long established knowings of how to work with particular novel phenomena and their associated functionalities. This necessary craft needs to be cultured slowly over decades in local settings with steady funding and encouragement. It is fed by universities; and it localizes because it tends to be shared at any time in any novel field of invention by small numbers of people confined to certain labs or to particular regions. But it does not last forever. Sooner or later knowings become mere knowledge and leak out; phenomena and principles become understood and are mined elsewhere; craft becomes codified into instruction and appears in textbooks; and people leave over time and spread their expertise geographically. Leadership in technology needs to be constantly renewed.

8. Conclusion

In this paper I have given an account of the structure of invention of radically novel technologies. Invention is not an event signaled by some striking breakthrough. It is process – usually a lengthy and untidy one – of linking a purpose with a principle (some generic use of an effect) that will satisfy it. This linkage stretches from the need itself to the base phenomenon that will be harnessed to meet it, through supporting solutions and sub-solutions and the grammars of each. The overall process may start anywhere along this chain: from a pressing need; from a novel phenomenon or the concept of its use; from the provision of some missing structure or element; or from knowledge or a piece of theory that enables these. The variations are many because the combinations of causal sequence are large, and the particulars of the problems to be solved differ. No two stories have the same plot, yet at bottom all share the same logical structure: all involve a conceptual linking of a purpose to a principle together with the resolution of the subproblems this causes. This linking defines a recursive process: it repeats until each subproblem resolves itself into one that can be physically dealt with. In the end the problem must be solved with pieces – components – that already exist (or pieces that can be created from ones that already exist). To invent something is to find it in what previously exists.

I do not want to claim that this structure applies to all domains of technology. Certainly there are two to which it does not. Modern combinatorial chemistry and synthetic biology create new functions (molecules or genetic regulatory pathways tasked to some purpose) by a process of random combination and subsequent testing. And as I remarked earlier, non-deliberate innovations such as trading arrangements or legal systems "emerge" via a social process of variation and selection. But outside these I believe the process applies to all modern deliberate purposed radical innovation. It also extends to much if not all pre-modern and prehistoric innovation. In the typical case (think of the origin of ore smelting, for example, which must have occurred multiply) an effect is observed, often accidentally, and rendered useful in some working form for some particular need; this is the phenomenon-initiated pattern I discussed above. Of

[8] For an analysis of technological revolutions see Perez (2002).

course, a great deal of variation and selection then follow as the technique in question is gradually improved.

To what extent is the argument above new? Certainly parts of the process I have described occur in earlier theories of radical innovation. But we can now see a distinction between invention induced by need versus induced by the discovery of a novel phenomenon. We can see how the various component stages of invention can be combined differently, explaining why the process varies from case to case. And we can see that invention has an understandable structure that follows from the properties of technology itself: that technologies are combinations; that they are phenomenon-based; and that their architecture is recursive. My aim throughout has been to show that invention has a logic – a systematic structure – albeit one that varies from case to case, that stems from properties common to all technologies.

Above all, invention is a process of recursive problem solving. As such it calls for thought, deep and persistent thought. But this does not mean it is a purely rational process, for at its core it consists in mentally associating a particular need with an abstract architected form that can handle it and in doing this repeatedly at several levels. Such associations form within the subconscious, and therein still lies a mystery: of how subconscious thought can dive into the depths of a problem and surface eventually with a solution that can meet it. But at least this is a known mystery, general to thought and not peculiar to the invention of technology.

Acknowledgements

I thank Kenneth Arrow, Niamh Arthur, Martin Bell, Arnulf Grübler, Gregory Nemet, Wolfgang Polak, Nathan Rosenberg, Martin Shubik, Nick von Tunzelmann, and two anonymous referees for useful comments. Any errors are of course mine. An earlier version of this paper appeared as "The Logic of Invention," Santa Fe Institute WP 2005-12-045.

References

Aitken, H.G.J., 1985. The Continuous Wave. Princeton University Press, Princeton.

Arthur, W.B., 1989. Competing technologies, increasing returns, and lock-in by historical events. Economic Journal 99, 116–131.

Arthur, W.B., 1994. Increasing Returns and Path Dependence in the Economy. University of Michigan Press, Ann Arbor.

Arthur, W.B., Polak, W., 2006. The evolution of technology within a simple computer model. Complexity 11 (5), 23–31.

Arthur, W.B., 2007. The Nature of Technology: What it is and How it Evolves. The Free Press, New York.

Aston, F.W., 1922. Mass spectra and isotopes. Nobel Lecture, 12 December.

Badash, L., 1998. Scientists and the Development of Nuclear Weapons. Humanity Books, New York.

Basalla, G., 1988. The Evolution of Technology. Cambridge University Press, Cambridge.

Bijker, W., 1995. Of Bicycles, Bakelite, and Bulbs: Toward a Theory of Socio-technical Change. MIT Press, Cambridge, MA.

Brown, J.S., Duguid, P., 2000. The Social Life of Information. Harvard Business Press, Cambridge, MA.

Buderi, R., 1996. The Invention that Changed the World. Simon and Schuster, New York.

Burns, R., 1988. The background to the development of the cavity magnetron. In: Burns (Ed.), Radar Development to 1945. Peter Peregrenus, London.

Callick, E.B., 1990. Metres to Microwaves: British Development of Active Components for Radar Systems 1937 to 1944. Peter Peregrinus, London.

Cathcart, B., 2004. The Fly in the Cathedral: How a Group of Cambridge Scientists Won the International Race to Split the Atom. Farrar, Straus, and Giroux, New York.

Chain, E., 1971. Thirty Years of Penicillin Therapy. Proceedings of the Royal Society of London B 179, 293–319.

Clark, R.W., 1985. The Life of Ernst Chain: Penicillin and Beyond. St. Martin's Press, New York.

Constant, E.W., 1980. The Origins of the Turbojet Revolution. Johns Hopkins University Press, Baltimore.

David, P., 1975. Technical Choice, Innovation and Economic Growth. Cambridge University Press.

David, P., 1985. Clio and the economics of QWERTY. American Economic Review 72 (2), 332–337.

Dosi, G., 1988. Sources, procedures, and microeconomic effects of innovation. Journal of Economic Literature XXVI (September), 1120–1171.

Freeman, C., 1990. The Economics of Innovation. Edward Elgar, Aldershot, England.

Gilfillan, S.C., 1935. The Sociology of Invention. Follett, Cambridge, MA.

Hare, R., 1970. The Birth of Penicillin. George Allen and Unwin, London.

Hiltzik, M.A., 1999. Dealers of Lightning: Xerox PARC and the Dawn of the Computer Age. HarperBusiness, New York.

Jewkes, J., David, S., Richard, S, 1969. The Sources of Invention. Norton, New York.

Kaempffert, W., 1930. Invention and society. Reading with a Purpose Series, No. 56. American Library Assoc., Chicago.

Kaiser, W., 1999. What drives innovation in technology? Technology & Culture 21, 107–123.

Kevles, B.H., 1997. Naked to the Bone: Medical Imaging in the Twentieth Century. Rutgers University Press, New Brunswick, NJ.

Köhler, W., 1926. The Mentality of Apes. Harcourt, Brace, London.

Kragh, H., 2002. Quantum Generations. Princeton University Press, Princeton.

Landes, D.S., 1983. Revolution in Time. Belknap Press, Cambridge, MA.

Landes, D.S., 1989. Hand and mind in time measurement. In: Kranzberg, M., Elkana, Y., Tadmor, Z. (Eds.), Innovations at the Crossroads between Science and Technology. Neaman Press, Haifa.

Lane, D., Maxfield, R., 1997. Foresight, complexity, and strategy. In: Aurthur, W.B., Durlauf, S., Lane, D.A. (Eds.), The Economy as an Evolving Complex System II. Addison-Wesley, Reading.

Lawrence, E., 1951. The evolution of the cyclotron. Nobel Lecture.

Lax, E., 2005. The Mold in Dr. Florey's Coat. Henry Holt, New York.

McGee, D., 1995. The early sociology of invention. Technology and Culture 36, 4.
Mokyr, J., 2002. The Gifts of Athena. Princeton University Press, Princeton.
Mullis, K., 1999. Dancing Naked in the Mind Field. Vintage Books, New York.
Murmann, J.P., 2003. Knowledge and Competitive Advantage. Cambridge University Press, Cambridge, UK.
Ogburn, W.F., 1922. Social Change. Dell, New York, p. 1966.
Otis, C., 1997. Aircraft Gas Turbine Powerplants. Jeppesen Sanderson Inc., Englewood, Colorado.
Perez, C., 2002. Technological Revolutions and Financial Capital. Edward Elgar, Aldershot, UK.
Polanyi, M., 1967. The Tacit Dimension. Anchor Books, New York.
Rhodes, R., 1986. The Making of the Atomic Bomb. Simon and Schuster, New York.
Rosenberg, N., 1982. Inside the Black Box: Technology and Economics. Cambridge University Press.
Ruttan, V., 1961. Usher and Schumpeter on invention, innovation, and technological change. Quarterly Journal of Economics LXXV, 596–606.
Schmookler, J., 1966. Invention and Economic Growth. Harvard University Press, Cambridge, MA.
Schumpeter, J., 1911. Theory of Economic Development. Harvard Press, Cambridge, MA, p. 1934.
Starkweather, G., 1980. High-speed laser printing systems. In: Ross, M., Aronowitz, F. (Eds.), Laser Applications, 4. Academic Press, New York.
Starkweather, G., 1997. Laser printer retrospective. In: Proceedings of the IS&T's 50th Annual Conference: A Celebration of All Imaging, Cambridge, MA, p. 702.
St. Peter, J., 2002. History of Aircraft Gas Turbine Development in the US. ASME, New York.
Süsskind, C., 1988. Who Invented radar?" In: Burns (Ed.). Op cit. 506-12.
Süsskind, C., 1994. Radar as a study in simultaneous invention. In: Blumtritt, Petzold, Aspray (Eds.), Tracking the History of Radar. IEEE, Piscataway, NJ, pp. 237–245.
Thumm, M., 2001. Historical German contributions to physics and applications of electromagnetic oscillations and waves. Proceedings of the International Conference on Progress in Nonlinear Science, Nizhny Novgorod, Russia, Vol. II. Frontiers of Nonlinear Physics, 623–643.
Thurston, F R., 1889. A history of the growth of the steam engine.
Townes, C.H., 1999. How the Laser Happened. Oxford University Press, New York.
Usher, A.P., 1929. A History of Mechanical Inventions. Dover, New York, p. 1988.
Whittle, F., 1953. Jet: The Story of a Pioneer. Frederick Muller, London.
Wideröe, R., 1928. Über ein neues Prinzip zur Herstellung hoher Spannungen. Archiv für Elektrotechnik XXI, 386–405.
Williams, T.I., 1984. Howard Florey: Penicillin and After. Oxford University Press, London.
Wilson, E.J.N., 1998. History of accelerators, Lecture 1. CERN, Geneva.