Abstract
Digital goods are bitstrings, sequences of 0s and 1s, which have economic value. They are distinguished from other goods by five characteristics: digital goods are nonrival, infinitely expansible, discrete, aspatial, and recombinant. The New Economy is one where the economics of digital goods importantly influence aggregate economic performance. This Article considers such influences not by hypothesizing ad hoc inefficiencies that the New Economy can purport to resolve, but instead by beginning from an Arrow-Debreu perspective and asking how digital goods affect outcomes. This approach sheds light on why property rights on digital goods differ from property rights in general, guaranteeing neither appropriate incentives nor social efficiency; provides further insight into why Open Source Software is a successful model of innovation and development in digital goods industries; and helps explain how geographical clustering matters.

This paper was produced as part of the Centre's Technology and Growth Programme

Communications to: Danny Quah, Economics Department, LSE, Houghton Street, London WC2A 2AE; Tel: +44/0 20 7955.7535; Email: dq@econ.lse.ac.uk
(URL) http://econ.lse.ac.uk/staff/dquah/

Digital Goods and the New Economy

Danny Quah

March 2003

DIGITAL GOODS AND THE NEW ECONOMY
by
Danny Quah

GLOSSARY

Arrow-Debreu model The standard model of general economic
equilibrium where a complete set of markets is available in all
commodities, indexed as necessary by time and state of nature;
all consumers and firms take prices as given; and consumer pref-
erences and production technologies are convex. Behaviour is
competitive or perfect competition prevails when all markets
clear with consumers maximizing preferences and firms maxi-
mizing profits taking prices as given. A competitive equilibrium

is a price and quantity configuration under perfect competition. An allocation—an assignment of specific quantities of consumption and production to consumers and firms respectively—is efficient if no other feasible allocation makes someone strictly better off and no one worse off. In Arrow-Debreu models, competitive equilibria produce allocations that are efficient.

Bitstring A string of 0s and 1s. Examples of bitstrings are engineering blueprints, chemical formulas, DNA sequences, mathematical theorems, computer software, digital music and images, and videogames. Everything that can be stored in computer memory and transmitted over the Internet is a bitstring. From mathematical logic, every statement—and therefore many items of knowledge—can be encoded as a bitstring.

Convex A set is convex when it contains all points on the line segment joining any two points in the set. Preferences are convex when the collection of consumer bundles preferred to any given bundle is a convex set. A production technology is convex when, first, it shows nonincreasing returns and, second, the collection of factor inputs that produce at least a given amount of output is a convex set.

Endogenous growth theory A body of economic theory directed at explaining why and how economies grow. Most endogenous growth theory focuses on technological progress and human capital accumulation but parts of it also concern the growth effects of ad hoc postulated inefficiencies—political vested interests, large fixed-cost barriers to high-productivity economic activity, ethno-linguistic or religious fractionalization, corruption, weak corporate and political governance, and so on.

ICT Information and communications technology. Since this includes computer hardware among other things, not all of ICT is just bitstrings.

Increasing returns A production technology shows increasing returns to scale or simply increasing returns when an equipropor-

tional increase in factor inputs results in a greater than proportional increase in output. Increasing returns technologies are not convex. Under constant returns to scale, an equiproportional increase in factor inputs results in an exactly proportional increase in output; under decreasing returns to scale, an equiproportional increase in factor inputs results in less than a proportional increase in output.

Open Source Software Computer software where the source code is made available for users to read, modify, improve, use, and redistribute. This contrasts with that where the source code is a legally-enforced proprietary secret, with only the executable machinecode image licensed for use. Leading instances of Open Source Software include `GNU/Linux` (a `Unix`-compatible operating system plus numerous applications and software tools) and `Apache` (webserver software).

Productivity paradox The puzzle that from the 1970s onwards, massive investment in ICT did not appear to improve substantially many economies' measured productivity.

R&D Research and development.

Welfare Economics, Fundamental Theorems The First Fundamental Theorem of Welfare Economics asserts that, under weak conditions, decentralized price-taking produces an efficient allocation. This is an *invisible hand* result: Uncoordinated actions by consumers and firms, individually responding only to the prices that each observes, not to any centralized, society-wide considerations, nevertheless produce an outcome that is efficient—i.e., the outcome is socially desirable, not just individually so. The Second Fundamental Theorem of Welfare Economics asserts that, under weak conditions, any efficient allocation can be attained using only decentralized price-taking behavior in consumers and firms, provided that society's resources are first appropriately redistributed.

Digital goods are bitstrings, sequences of 0s and 1s, that have economic value. They are distinguished from other goods by five characteristics: digital goods are nonrival, infinitely expansible, discrete, aspatial, and recombinant. The New Economy is one where the economics of digital goods importantly influence aggregate economic performance. This Article considers such influences not by hypothesizing ad hoc inefficiencies that the New Economy can purport to resolve, but instead by beginning from an Arrow-Debreu perspective and asking how digital goods affect outcomes. This approach sheds light on why property rights on digital goods differ from property rights in general, guaranteeing neither appropriate incentives nor social efficiency; provides further insight into why Open Source Software is a successful model of innovation and development in digital goods industries; and helps explain how geographical clustering matters.

I.   Introduction

As documented elsewhere in this Handbook (and attested to by journalistic frenzy in the late 1990s' dotcom boom) the New Economy means different things to different observers. Possible dimensions to the New Economy range from e-commerce, e-government, the Internet, the productivity paradox, knowledge-intensive work, social mass-mobilization, and globalization, all the way through auction proliferation, electronic payment systems, venture capital financing saturation, and business restructuring. In less guarded moments, popular conception held that with the New Economy, inflation might be forever conquered, explosive income growth might be hereafter the norm, and stock markets be always stratospheric.

Whether those possibilities are real, now or in future, is not this Article's concern. Rather than studying the New Economy—whatever it might mean—by beginning from ad hoc implicit economic frictions that the New Economy can then purport to overcome, this Article adopts the opposite attack. It takes a background perspective of markets in perfectly competitive Arrow-Debreu equilibrium, and asks, What is distinctive about the New Economy in general or

digital goods in particular that could affect economic performance?

This strategy, in the current author's view, preserves analytical rigor and discipline. But, perhaps more important, since many observers consider the ideal of zero transaction cost, instantaneously buyer/seller-matched, friction-free, transparent, perfect-information markets to be the end result of the New Economy in any case, studying what happens at that limit point—i.e., what textbook economics has always assumed in the Arrow-Debreu model—might yield more enduring insight than will studying the hypothesized transition towards it.

This Article provides a definition of digital goods in the New Economy and describes a number of scientific, social, and commercial developments relating to that definition. The Article considers both traditional and recent formulations of the economics of digital goods—ideas; knowledge and economic growth; intellectual property; and nonrivalry, infinite expansiblity, discreteness, aspatiality (or, weightlessness and spacelessness), and recombination. Of course, not all these conceptualizations were designed originally with an eye to what I call digital goods in this Article, but the underlying economic principles nevertheless apply. This Article, therefore, takes the economics of austere high science, technology, and R&D to apply with equal force to videogames, movies, and pop music, as to biotechnology and computer software. In this framework, some digital goods and some parts of the New Economy have a lot to do with knowledge, skills, and productivity; others, hardly at all.

The discussion to follow considers, among other things, the difference between nonrivalry and infinite expansibility. Traditionally, economists have taken these two properties to be equivalent; indeed, for many interesting questions they should be treated thus. However, in recent work on pricing ideas without the artifact of intellectual property rights, the distinction between nonrivalry and infinite expansibility matters. This Article explains that difference.

Theories of increasing returns and network externalities apply to digital goods as a special case. Consequently, digital goods and the New Economy can be expected to display behavior such as cumulative causation, path dependence, production and consumption spillovers,

and what Sherwin Rosen labeled the economics of superstars. But, even if particularly pronounced for digital goods, such predictions are not special to them. Indeed, as interpreted by the contributors to those literatures, their analyses apply to a wide range of economic activity, including traditional manufacturing. Since those ideas are sufficiently rich and intricate to merit detailed exposition elsewhere, this Article steers clear of them and focuses instead on what is unique to digital goods in the New Economy.

A   What is a digital good in the New Economy? What isn't?

A *digital good* is a payoff-relevant bitstring, i.e., a sequence of binary digits, 0s and 1s, that affects the utility of or payoff to some individual in the economy. Easiest is to think of a digital good as a *recipe*: Encoded in the digital good (and, indeed, identical with it) is a set of economically valuable instructions. The phrasing allows digital goods to be consumed and to be produced; they are not just technologies to improve productivity on the supply side of an economy.

Any copy of a digital good is the good itself. There is no distinction between an original and a copy. No one holding a digital good relinquishes possession of it when yet others gain it; no one acquires a digital good by necessarily confiscating it from someone else. Indeed, the first owner will be unaware altogether of additional acquisition not of copies—which would not be at all unusual—but of the good itself.

Ideas and knowledge, computer software, visual images, music, databases, videogames, blueprints, recipes, DNA sequences, codified messages, and so on are all digital goods. Are there visual images that are *not* digital goods? Yes, examples include works of art for which the smell of the canvas, the texture of the oilpaint, or the perceived brushstroke by a long-dead artist, distinguish the original from its copies.

In this definition, a useful distinction is between digital goods that are *robust* and those that are *fragile*. If the economic value of the

good is unchanged when a sufficiently small but positive fraction of the bitstring is randomly removed or re-assigned (i.e., the bitstring is contaminated), then say that digital good is robust. Otherwise, say the digital good is fragile. Typical lists of instructions that are the machinecode for a piece of computer software will refuse to execute when contaminated in the slightest, and so are fragile. Similarly, vector encodings of images—lists of abstract instructions—are fragile. Digital music recordings and bitmapped digital images, on the other hand, are robust: Indeed, that is how compression techniques such as JPEG and MP3 encodings work, producing shorter bitstrings with the same economic value as the original. Such compression—permanently changing the data and shedding the ability to re-create the uncontaminated original—differ from so-called lossless compression, where the original can be recovered perfectly from a compressed image, even though the latter is a strictly shorter sequence of 0s and 1s. In lossless schemes, the compressed image is generated deterministically, not randomly, from the original. Given current state of knowledge in genetics—although some recent research disputes this—contamination confined to the 97% of gene sequences in so-called "junk DNA" in human cells produces no change in the effectiveness of human DNA in coding and manufacturing proteins. However, contamination occurring over the remaining 3% results in mutation. Thus, we might usefully consider human DNA—compared to other digital goods—to display a sliding scale of fragility. Almost all of what follows in this Article applies simultaneously to both fragile and robust digital goods, but the distinction sometimes matters and so is useful to keep in mind.

Innovation, in this analysis, is the instantiation, i.e., the first creation, of a digital good. The New Economy, then, is an economy where digital goods figure prominently in determining aggregate economic outcomes—innovation, production, and consumption.

Economics has traditionally viewed digital goods as ideas, i.e., scientific knowledge, engineering blueprints, and technological innovation. That historical identification makes it natural to associate digital goods with improvements on the production or supply side of the economy. In that view, the New Economy is a knowledge-driven
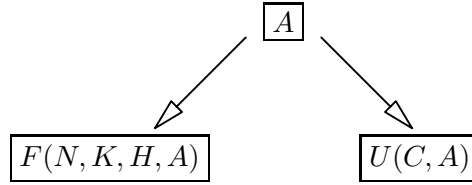
$$\boxed{A}$$

$$\boxed{F(N,K,H,A)} \qquad \boxed{U(C,A)}$$

Figure 1: Let A denote digital goods. The left arm in the Figure points to firms' production functions F; the right, consumers' utility U. Potentially different A's enter production and consumption. What matters is that they all share the same essential economic properties. In production function F, symbol N denotes labor; symbols K and H, physical and human capital, respectively. In utility function U, symbol C denotes ordinary consumption goods. The Figure illustrates, therefore, the traditional view, that digital goods contribute to production on the supply side of an economy, as well as the newer view taken in this Article, that digital goods can also contribute directly to utility from their consumption by final consumers on the demand side.

economy, with productivity rising when technology advances due to knowledge accumulation.

To emphasize this historical association, turn to Fig. 1 and call digital goods A, the traditional symbol denoting technology in economic growth theory. The left side of Fig. 1 shows a production function F, mapping to total output the state of technology A and factor inputs (N, K, H), labor and physical and human capital. In this stylization, improvements in A raise productivity and drive economic growth. Obvious examples of such A include engineering blueprints, chemical formulas, and industrial innovations, i.e., intellectual property protected by either formal institutions such as patents or informal ones such as trade secrets. Call items of knowledge that can be so encoded *codifiable*. Call *tacit* all other items of knowledge—since in the framework of Fig. 1 these are embodied in economic agents, we might as well call this human capital.

(This terminology, although sufficient for the purposes of this Article, does serious injustice to long-established literatures in eco-

nomic history, technology, epistemology, and sociology, among others. Robin Cowan, Paul David, and Dominique Foray provide a useful summary of some of these disparate strands of thinking across the social sciences more generally.)

The right side of Fig. 1 adds the possibility that digital goods directly affect consumers' utility. Compared to traditional views on knowledge in economic growth, this might at first appear peculiar. However, a moment's reflection readily provides examples: videogames, digital images and music (i.e., new media and entertainment), computer software, biotechnology, and significant portions of the telecommunications industry and the Internet. To be clear, genetically-modified frost-resistant tomatoes, say, should certainly be expected to influence productivity; but picture messaging, music, and games on mobile telephone handsets—features driving new-generation telephony—less obviously so. However, these last and numerous other examples like them comprise large and growing shares of some modern economies. Indeed, considering the impact of this part of the New Economy, the real paradox would be if supply-side productivity measurements *were* affected!

The two arms in Fig. 1 flag perspectives that are both analytical and empirical. Just as the economics of digital goods in production differ from that in consumption, so too might measurement of the New Economy usefully consider developments on both the demand side and the supply side. Empirical analysis of the New Economy might profitably study not only whether computers raise labor productivity, say, but also how risk allocation and consumption patterns, political organization and mobilization, and so on are evolving with increasing computer and Internet proliferation.

Notably omitted from Fig. 1 are those considerations in the traditional economics of information and uncertainty—risk and agency, moral hazard and adverse selection, signalling, and strategic behavior under asymmetric and imperfect information. (See, for instance, the textbook presentation in Jack Hirshleifer and John Riley.) While common parlance holds that a growing New Economy entails the rising importance of information in economic activity, that description is usefully distinguished from how economics has traditionally taken

–9–

information to matter. The economics of valuing and disseminating, say, an MP3 music file differs from the economics of moral hazard in sharecropping or of adverse selection in insurance.

How are biotechnology products also digital goods? Biological development entails manipulating genetic material. This last is just a DNA sequence, i.e., a string of subunits comprising one of four nitrogen compounds, Adenosine, Cytosine, Thymine, and Guanine (or, the letters A, C, T, G). DNA sequences translate into sequences of 0s and 1s, and thus are bitstrings that code for—bear information allowing selected cell sites to create—different proteins in plants and animals.

While useful, making this last specific connection explicit potentially disguises how the digitization idea applies more generally. It is a commonplace that numbers are digital, either directly or when translated into binary representation. Increasing intrusion of computers and the Internet into everyday activity has made familiar the idea that software, music, images, and so on are digital.

What is perhaps less well-known but useful to clarify in an article on digital goods is how digitization—the identification of the humble bitstring with many objects in the modern world—might be one of the surprise (and implicit) grand unifying themes of 20th-century science and economic progress. The next section briefly describes some of that intellectual background. While not absolutely essential to the remainder of this Article, it usefully provides a broader framework to the discussion.

B Eating 1s and 0s

Readers will be familiar with the idea that computer games, digital music, and video images are strings of 1s and 0s stored on a computer hard disk or an Internet webserver. Some take such observations to mean digitization is confined to superficial and frivolous leisure activity.

That identification misleads. Bitstrings that are economic goods, and digitization more generally, have a long and venerable intellectual

–10–

history in science. Appreciating that background provides perspective on the discussion to follow.

In Fig. 1 production function F implicitly gives, among other things, elasticities of substitution across factor inputs. The specification F is of course general but that same generality hides special and interesting properties that arise when factor inputs include bitstrings.

Physical capital K is naturally taken to denote machines or hardware. Ordinary machines provide a fixed, finite range of actions: attach a door onto hinges in an automobile assembly line, shift a given volume of soil, decant a specific quantity of chemical reactant. While a microwave oven can act upon a whole range of different objects, it performs only one significant operation: agitate molecules. Similarly, while many different objects can be placed inside it, a refrigerator does only one thing of note: cool objects. Numerical-control machine tools are a bit more flexible, but their range of operations too is delineated ahead of time. In all these cases, the original design specifies, once and for all, the set of possible actions. Hardware, unaugmented, can execute no operation that surprises its designer. As a rule, the more modern is such a machine, the more specialized and tightly-structured it becomes, and the more restricted and fixed in advance its range of actions.

By contrast, bitstrings drive towards universal and self-modifying operation. The Turing machine, conceptualized by Alan Turing in the mid-1930s to determine the decideability of mathematics as a logical system of axioms, made precise how a small fixed set of instructions—read one token of input from a stream; write one token onto the same stream; move to the previous token or to the next token—together with a finite set of states or configurations of the machine could produce effectively all conceivable computable outputs. Put differently, finite hardware becomes general-purpose when a (finite) Turing machine is attached to it and then fed an appropriate bit-string sequence of instructions. (Notice that this general-purpose nature differs from that more commonly understood in economics under "general-purpose technologies", which concerns instead how a given technology can simultaneously affect many different sectors of an economy.)

–11–

After the Second World War, John von Neumann's and Alan Turing's design of stored-program machines permitted seamless intermingling of instructions and data—both bitstrings—in computer core memory, and therefore extended operation to instruction sets that could easily and quickly self-modify. This development raised the possibility that machines can, for all practical economic purposes, adapt and learn—and thus, in language used earlier in this Article, surprise their designer. Obviously, the current discussion intends to raise no deep controversial issues in epistemology or psychology; it only observes that self-modification can have interesting and surprising *economic* implications, for instance in how self-modifying bitstrings can alter the production process. Self-modification applied to bitstring pairs, for example, characterizes what Martin Weitzman calls recombinant growth, described later in this Article. Arising naturally from the general point on self-modification are the important notions of evolutionary self-organization and emergence, that have been studied in complexity theory, although relatively unexplored still in economics.

Logically prior to these developments is the invention of Gödel numbers—integer representations of mathematical proofs and statements, first as logic symbol-strings and then as prime number power multiples, thereby mapping axiomatic systems in mathematics into bitstrings. In the 1930s Kurt Gödel and Alan Turing had sought only to study the logical foundations of mathematics in the Hilbert program—as had many illustrious mathematicians before them—but the encoding tools they developed ended up useful also for the entirely new area of information and communications technology. Intimately related, number theory in mathematics—previously thought to have zero practical significance—now sees commercial application in economically-valuable bitstring manipulation: in intellectual property protection and cryptography. Such digital goods, therefore, are used in preserving the integrity and authenticity of exchange in financial assets and yet other digital goods.

II.   Some economics of digital goods

This Article has chosen to formalize digital goods as bitstrings, rather than leaving them defined as just information or knowledge, terms where namespaces are already overloaded. The selected formalization allows greater precision—and, in light of the discussion in section I.B, follows a sound scientific tradition—so that the economic implications of digital goods can here be more easily developed.

Digital goods, by their nature, have five properties central to the discussion in this Article. Digital goods are nonrival, infinitely expansible, discrete, aspatial, and recombinant. Discuss each of these properties in turn and then consider their implications.

A good is *nonrival* when its use by one agent does not degrade its usefulness to any other agent. Thus, ideas, mathematical theorems, videogames, engineering blueprints, computer software, cookery recipes, the decimal expansion of π, gene sequences, and so on are nonrival. By contrast, food is distinctly rival: consumption renders it immediately no longer existent.

[Excludability, often discussed in the same breath as nonrivalry, is ancillary—it is not primitive to digital goods but instead follows from a hypothesized enforcement mechanism protecting them. Enforcement mechanisms can be legal or technological or both. Thus, intellectual property rights, discussed at greater length later in this Article, are legal mechanisms disallowing certain specific operations on digital goods. Encryption (a digital good itself) is a technical device that can be used, similarly, to constrain how digital goods are used. Such encryption might fall within the law or outside it, or simply be indifferent to the law altogether. Excludability, therefore, can arise from the law or from technology or from both, but it is not itself intrinsic to digital goods. This Article will discuss excludability no further.]

A good is *infinitely expansible* when its quantity can be made arbitrarily large arbitrarily quickly at no cost. Infinite expansibility is why media companies fear that digital music and images—costly for them to produce but distributed freely over the Internet—will proliferate without bound.

–13–

Historical roots for the concepts of nonrivalry and infinite expansibility go back at least to the early 19th century, where they are associated most with the writings of Thomas Jefferson. Widely cited by scholars and observers is Jefferson's 13 August 1813 letter to Isaac McPherson:

> "If nature has made any one thing less susceptible than all others of exclusive property, it is the action of the thinking power called an idea, which an individual may exclusively possess as long as he keeps it to himself; but the moment it is divulged, it forces itself into the possession of everyone, and the receiver cannot dispossess himself of it. Its peculiar character, too, is that no one possesses the less, because every other possesses the whole of it. He who receives an idea from me, receives instruction himself without lessening mine; as he who lights his taper at mine, receives light without darkening me. That ideas should freely spread from one to another over the globe, for the moral and mutual instruction of man, and improvement of his condition, seems to have been peculiarly and benevolently designed by nature, when she made them, like fire, expansible over all space, without lessening their density at any point, and like the air in which we breathe, move, and have our physical being ... ."

Jefferson's forceful writing here has been used, variously, either to defend strong intellectual property rights—because Jefferson was one of the founders of the US Patents Office, and the passage above can be read as calling for such an institution—or, more directly and typically, to justify doing away with intellectual property rights altogether.

In the quoted passage, Jefferson describes infinite expansibility when he discusses how ideas (or bitstrings, in our terminology) are such that "no one possesses the less, because every other possesses the whole" and how ideas are by nature "expansible over all space, without lessening their density at any point". On the other hand, Jefferson describes nonrivalry in how "he who receives an idea from me,

receives instruction himself without lessening mine; as he who lights his taper at mine, receives light without darkening me." The first of these descriptions is most easily visualized as making arbitrarily many copies of a bitstring; the second, as using the copies extant of a given bitstring—the number of copies being fixed and finite—without drawing down the usefulness to any other user. This Article will return to the point later but the reader can, for now, take away that infinite expansibility implies nonrivalry, but nonrivalry can hold with or without infinite expansibility. Nonrivalry and infinite expansibility, while close, nevertheless differ.

Although nonrivalry had been used earlier in other areas of economics, it and infinite expansibility have attracted attention from macroeconomists primarily only since the 1990s. Paul David in 1992 first used the term "infinite expansibility" in analyzing knowledge dynamics in economic growth and development, relating it to Thomas Jefferson's writings in particular. In 1990 Paul Romer, studying endogenous technology and economic growth, put nonrivalry explicitly to the fore in his analysis. (See also Charles Jones's excellent textbook presentation of economic growth.) All this work takes knowledge and ideas—specific instances of digital goods—to be factor inputs in production; this work concerns the left side of Fig. 1.

Economists have generally viewed interchangeable the two concepts, nonrivalry and infinite expansibility. Each implies increasing returns and therefore nonconvexity, and each to the other had seemed logically equivalent. In that view, digital goods displaying increasing returns in production matters for two reasons. Were returns to scale constant rather than increasing, paying each factor of production its marginal product—as would occur under perfect competition—would exactly exhaust total output. However, with increasing returns, this precise adding-up no longer holds: the sum total of competitively-determined factor payments exceeds output, so that using only perfect competition in factor markets to organize production is infeasible. Factor inputs are so productive at the margin that their market-based compensation would call for paying out more than is actually produced in total. Second—related but different—when output is a digital good, production beyond the first instantiation occurs at

–15–

zero marginal cost so that perfect competition in the output market would give Arrow-Debreu equilibrium price equal to zero. But with zero price and costly instantiation upfront, no one would see private incentive to create the first instance of the digital good. Therefore, conventional markets for digital goods are doubly expected to fail.

For certain analyses, however, it matters that nonrivalry and infinite expansibility do not coincide—close though they might be. Nonrivalry describes a restriction on marginal utility or marginal productivity. Expansibility, on the other hand, describes a restriction on quantity available to society at zero marginal cost over a specific timespan.

An instance of nonrivalry in music is my enjoying a piece of opera, simultaneously as does another consumer in the same venue but with neither of us knowing the other to be present. One consumer's enjoyment is invariant to the other's. [To be clear, since such descriptions sometimes quickly and carelessly veer towards more intricate ideas, note explicitly here that what has been just described is much simpler than, say, network externalities or consumption spillovers.] This statement on utility in simultaneous consumption is blind to whether the opera is being performed live or taken off a high-quality digital recording. Of course, the utility level might vary—many consumers genuinely enjoy live opera in ways that recorded opera cannot achieve—but not the fact itself of joint and simultaneous enjoyment. Since live opera is a once-only event ("each time, it's different"), the live performance scenario displays nonrivalry but not infinite expansibility. Or, nonrivalry is possible without infinite expansibility.

Infinite expansibility, on the other hand, always generates nonrivalry. Being able to make arbitrarily many copies of a digital good means that everyone can have their own copy, with every copy of a digital good being again exactly an original. Consumers can, therefore, enjoy a copy or producers can use a copy, without drawing down the good held by any other consumer or producer. Infinite expansibility generates nonrivalry by brute-force copying, flooding the market. Nonrivalry without infinite expansibility is more subtle, and the good concerned can remain bounded in quantity below the extent of the market.

This distinction has implications for Arrow-Debreu equilibrium. With infinite expansibility, the Arrow-Debreu price equals zero, the marginal cost of reproduction in the digital good. But if the first copy of the digital good uses up resources in its instantiation, then a zero price results in market failure: A socially worthwhile good is left unproduced in equilibrium. By contrast, with nonrivalry but only finite expansibility, Arrow-Debreu prices remain positive and, under appropriate conditions, can produce a socially efficient outcome. (Michele Boldrin and David Levine and the current author have studied these circumstances.)

The discussion leads naturally to the third special characteristic of digital goods, namely that they are (initially) discrete. Of course, copies always come in integer amounts, so that, trivially, digital goods are always discrete—the quantity of a digital good can only be 0, 1, 2, 3, ... But that is unimportant; instead, what will turn out to matter is that digital goods only ever instantiate to quantity 1.

An alternative description of discreteness is that digital goods show indivisibility. A half-baked idea can be worse than no idea at all. This is most obvious for fragile digital goods. The first half of the string of 1s and 0s constituting a computer program will not execute half or indeed any positive fraction of the program's intended task and might instead damage the computer hardware; an incomplete string of exons (the functioning, non-junk DNA) in a gene sequence will not express part of a protein molecule; the first few lines of the proof of a deep mathematical result do not constitute a proof of anything. Even robust digital goods, by the definition, remain payoff-equivalent only for small contamination. Making a fractional copy rather than a whole one, where the fraction is distant from 1, wil destroy that particular instance of the digital good.

The importance of discreteness or indivisibility lies in the economics surrounding the digital good's instantiation. Since individual consumers or producers always use a whole copy of a digital good, it is less significant that fractional copies are not available or useful afterwards. When, however, the first copy of a digital good to be instantiated requires considerable investment—completing the first copy of a new computer operating system; writing down the first proof of a

–17–

deep mathematical theorem; isolating the first DNA sequence that codes against the gene mutation and runaway cell growth in cancer— then this initialization cost drives a wedge between what is socially optimal and what perfectly competitive markets can deliver, even under finite expansibility. (To mix metaphors, that wedge is the size of a Dupuit triangle.)

The three features described thus far—nonrivalry, infinite expansibility, initial discreteness—are all special cases of increasing returns or nonconvexities, and so naturally share certain common implications for market equilibrium. Arrow-Debreu equilibria need no longer exist; but even when they do, they need no longer be socially efficient.

But saying simply increasing returns or nonconvexities misses important subtleties. As just discussed, these three features of digital goods have critically different implications. Infinite expansibility can be usefully replaced by a finite approximation, e.g., where the number of copies can get arbitrarily large only gradually, or only with some small but positive cost. This captures how even Internet dissemination of computer software or digital music is constrained by finite bandwidths of networks and gateways; how the spread of ideas and knowledge is slowed by social institutions and bounded human capacities; and so on. With expansibility finite, nonrivalry alone presents problems for neither Arrow-Debreu pricing nor social efficiency in perfectly competitive markets. Those conclusions can remain even when the degree of finite expansibility grows without bound, i.e., when digital goods approach infinite expansibility. This limiting result breaks down, though, if market trading is allowed to occur continuously, rather than only at discrete time intervals, whereupon with infinite expansibility market failure then again applies. Finally, for digital goods, indivisibility need present no special difficulties for the existence of Arrow-Debreu equilibrium. However, if the indivisibility exceeds a minimum threshold scale—which depends on consumer tastes, among other things—then Arrow-Debreu equilibrium will be inefficient.

The fourth feature: Digital goods are aspatial; they are both nowhere and everywhere at the same time. Just as any copy is the original for a digital good, so too communication of a digital good

is its transportation and distribution. Whether or not in practice all digital goods now, uniformly and immediately, "freely spread from one to another over the globe", obviously it is their nature to do so. [While, under certain interpretations, aspatiality might be considered only another instance of increasing returns, this author's view is that, again, doing so gives no special insight.]

Finally, the fifth property: Digital goods are recombinant. By this I mean they are cumulative and emergent—new digital goods that arise from merging antecedents have features absent from the original, parent digital goods. (The terminology derives from recombinant DNA and genetic recombination.) Digital goods generate new digital goods in ways unavailable to, say, combining ordinary public goods like clean air, national defense, or a lighthouse. Martin Weitzman has shown that the combinatorial properties in merging pairs of bitstrings imply growth of new ideas exceeding any fixed, finite exponential rates.

III.   Implications and puzzles

Digital goods' distinctive features—that they are nonrival, infinitely expansible, discrete, aspatial, and recombinant—help explain important observations in the New Economy. At the same time, they raise several significant puzzles.

A   Intellectual property

Infinite expansibility in digital goods generates a tension between ex post efficiency and ex ante incentive. To see this, recall as described earlier how if a digital good were traded in competitive markets, its price would equal marginal cost, namely zero. Thus, as long as someone continues to value the digital good, no matter how low that positive valuation at the margin, producers should and will pump out more and more of the good, up until market saturation. Competitive markets drive the price on the digital good to zero. That is the ex post efficient outcome.

This outcome can obtain in different ways. One explicit mechanism enforcing it is to permit early purchasers of the digital good to make copies at marginal cost so that those purchasers can then freely compete with the original owner, i.e., to disallow the trade restrictions embodied in intellectual property rights (IPRs). In an Arrow-Debreu environment, such exchange of digital goods, peeled off at zero marginal cost, is neither theft nor piracy. Instead, the zero price simply shows markets working.

But, however that ex post efficient outcome is achieved, the stream of rents thus generated fails to incentivize sufficient instantiation of such goods. Society sees too little innovation in Arrow-Debreu equilibrium. That the equilibrium competitive price is exactly zero is inessential to the argument, critical is only that the reproduction marginal cost is sufficiently low.

As already suggested implicitly in the language used earlier, one way to encourage innovation is to restrict the trade in digital goods through assigning monopoly rights in distribution. IPRs—patent, copyright, trademark protection, and others—constitute examples of such arrangements. The intellectual property (IP) owner can then monopolistically price the digital good to maximize profits, without that price falling to zero from competition. For this, early purchasers of the digital good must be prevented, by law, from making copies of the good for resale, as would occur under perfectly competitive markets. To repeat, in this description, absent IPRs, such untramelled copying would be competition, not theft.

While such IPR protection raises the rental stream accruing to the owner of the digital good and therefore encourages innovation, it also inflicts an inefficiency on society. Dissemination of the digital good becomes curtailed relative to the efficient market-saturating outcome. Compared to what the production technology implies, there is too little use of the digital good. IPR protection, ex post, delays the socially beneficial widespread application of already-instantiated digital goods.

These difficulties become particularly pronounced with pharmaceuticals and life-critical medication. The marginal cost of running off extra copies of a medication—reproducing the digital encoding of

the chemical information into insignificant physical material—is tiny, whereas the instantiation costs of a first working copy can run into the billions of dollars. Even if initial discovery were entirely serendipitous, the costs of clinical trials and government certification can still be substantial. On the other hand, for medication that successfully and uniquely treats an ailment (otherwise implying certain death to hundreds of millions of impoverished, afflicted Africans—as for AIDS, say), the ex post social cost of the IPR-based market-restricted outcome can be considerable.

The production and distribution of digital goods, therefore, display a unique social tension. IPRs differ profoundly from ordinary property rights. This will be familiar to most economists but, in less rigorous discussion, many observers seem to accord IPRs the same economic status as Arrow-Debreu based property rights. Those observers suggest that strong IPR protection is always and everywhere central to sound economic performance—by misguided analogy to ordinary property rights. The reasoning in this Article and elsewhere shows that such a view is untenable. Sometimes strong IPRs are appropriate; sometimes they are not. What is optimal to guide the production and distribution of intellectual assets or digital goods is, in general, quite subtle. The tenebrous advantages to IPR schemes are quite distant from the unalloyed benefits that derive from assigning ordinary property rights in production and exchange.

Thomas Jefferson's writings are again relevant here. In his 1813 letter to Isaac McPherson, in segments not as widely-known, Jefferson brackets the quotation already given with:

> It has been pretended by some, (and in England especially,) that inventors have a natural and exclusive right to their inventions, and not merely for their own lives, but inheritable to their heirs.

and then, later, continues with:

> Accordingly, it is a fact, as far as I am informed, that England was, until we copied her, the only country on

> earth which ever, by a general law, gave a legal right to
> the exclusive use of an idea. In some other countries it
> is sometimes done, in a great case, and by a special and
> personal act, but, generally speaking, other nations have
> thought that these monopolies produce more embarrass-
> ment than advantage to society; and it may be observed
> that the nations which refuse monopolies of invention, are
> as fruitful as England in new and useful devices.

Thus, although Jefferson was one of the founders of the US Patent
Office, and between 1790 and 1793 was one of its most scrupulous
examiners, Jefferson's views on IPRs were finely balanced. In the
passage, Jefferson showed he considered IP monopolies to be unde-
sirable, other things equal, and he asserted technical progress to be
achievable without IP protection. However, at the same time, Jeffer-
son well recognized the pragmatic difficulties of one extreme or the
other, for he concluded the letter with:

> Considering the exclusive right to invention as given not
> of natural right, but for the benefit of society, I know well
> the difficulty of drawing a line between the things which
> are worth to the public the embarrassment of an exclusive
> patent, and those which are not ….

Ultimately, the phrasing that appears in the US Constitution as Ar-
ticle 1 Section 8 Clause 8 allows Congress:

> To promote the Progress of Science and useful Arts, by
> securing for limited Times to Authors and Inventors the
> exclusive Right to their respective Writings and Discover-
> ies.

Thus, consistent with the argument in this Article, neither the US
Constitution nor one of its most thoughtful, articulate authors con-
sidered strong, unlimited-extent IPRs to hold unambiguous advan-
tage.

Modern IPR law takes among its goals not just protection but also *disclosure* of ideas. Indeed, in both 14th-century England and 18th-century US, domestic patents were awarded primarily to incentivize transferring technology already developed abroad into domestic usage. Patents thus served both to protect and to disseminate ideas. In practice now, however, at least for digital goods the disclosure function appears little emphasized by either IP law's proponents or critics. If the law achieved only disclosure and not at all protection, then it would be only a cipher relative to the economic forces discussed earlier in this Article.

To be clear, most IPR law was not designed with an eye to the economics of digital goods. Historically, patents have covered primarily industrial application, i.e., the mechanical or chemical transformation of physical material. To be patented, an idea had to be demonstrated to be novel relative to the state of the art, to involve a non-obvious invention, and to be capable of exploitation in industry. In that traditional view, mathematical algorithms, say in cryptography, would not be patentable since they had no apparent industrial application.

Copyright, by contrast, has traditionally targeted mostly literature, music, and works of art. It has mostly insisted artistic works be fixed in some tangible medium of expression. Unlike for patents, where stringent tests need to be met before a patent can be awarded, copyright protection is automatic. No formal registration is required. Not only can the copyright symbol © be freely placed on any work of art by that work's creator, in Berne Convention law such placement is not even necessary for that product to attract copyright protection. Thus, for instance, in the US or the UK every original piece of homework that a student hands in is automatically awarded copyright protection by law, whether or not marked ©.

In England, from at least the middle of the 16th century, copyright license had been used to provide monopoly income to stationers and other publishers. Nominally, the 1709 Statute of Anne re-assigned those rights to authors, but in practice power remained with publishers and booksellers, who continued this commercial focus on copyrighted material. As Thomas Jefferson had observed, however, England was unusual in this emphasis on economic exploitation of such

–23–

intellectual property. Elsewhere in the world—France in particular—copyright was viewed not as a means to guarantee an incentivizing revenue stream but instead as an expression of reverence to an author and as a moral right awarded to preserve artistic integrity and stylistic flair in an idea's implementation.

In all cases, however, copyright protection applies not to the underlying idea—e.g., the law of one price in economics homework assignments—but only to the idea's expression. The reasoning is that the raw idea is part of nature and therefore not created by any human artist. No issue of artistic integrity is involved in the idea itself. Thus, in this reading of IPR law, even as certain mathematical algorithms are unprotected under patent law if they don't have a ready industrial application, *no* mathematical algorithm can be protected by copyright law. No algorithm can be expressed with artist-specific flair; every algorithm must be part of the underlying state of nature. (This need not permit, however, directly reverse-engineering the implementation of a mathematical algorithm, embedded in hardware or software, for one could argue that that particular expression of the algorithm is protected by copyright.)

Generally then, when a digital good is covered by copyright, what is most valuable in it—the underlying idea—is unprotected. On the consumption side, a creator's flair potentially does matter, although the degree to which that happens will vary with the consumer and with the digital good. However, on the production side of an economy, an individual's implementation style for a critical algorithm is likely more distracting than it is substantive; at best, it is simply ignored. Thus, if the copyright owner is to capitalize on IP protection, some other means must be found to prevent the zero-price outcome. Since what *is* protected is the flair of expression or implementation, it is that dimension along which the IP owner and potential competitors will seek to distinguish themselves. But such competition is wasteful: while costly of resources, in the main it does little to improve the set of economically valuable bitstrings available to society. In computer software, this manifests in how the look and feel, the location of controls, the sets of menus holding commands, and so on, differ across products otherwise functionally equivalent. This makes such

products difficult to use, and builds in incompatibilities across the skillsets of different individuals. Socially beneficial, instead, would have been to keep exactly the same flair of implementation, but to develop new bitstrings that do more, i.e., have better algorithms in them. That, currently, most computer software is putatively covered under copyright might therefore be viewed, with considerable justification, to be makeshift shoehorning of digital goods into an ill-fitting and inappropriate older system of IPR protection.

Beyond this, digital goods need not be contained in a tangible medium of expression. Digital goods are economically valuable, their moral value irrelevant (at least in economic discussion)—thus, protection of digital goods under IPR law should be to generate a sufficient stream of economic rents, not only to preserve intellectual integrity in the abstract. By the same token, digital goods' protection under law should not be automatic but instead subject to rigorous scrutiny. One traditional, although not entirely accurate, description of the difference between patents and copyrights is that the former deals with machines, the latter with texts. Computer software and other digital goods erode the boundary between machines and texts. Thus, digital goods fit ill within protection frameworks afforded by both patents and copyrights, under current IP law.

We have seen that infinite expansibility—implying nonrivalry— calls for an institution beyond just Arrow-Debreu markets wherein to produce and exchange digital goods. Intellectual property rights restore the possibility of market-based exchange of digital goods. However, this veneer of reverting to Arrow-Debreu markets fails to restore their social efficiency implications. In even the simplest, most stylized environments, IPR protection on digital goods hardly ever produces socially desirable outcomes. Unsurprisingly, more complicated, more realistic situations imply only ever-escalating tensions between ex post efficiency and ex ante incentives in the workings of digital goods markets.

That property rights prevent a tragedy of the commons—over-exploiting a resource that no one agent has incentive to conserve—is a profoundly important insight. For ordinary goods, private property rights allow markets to emerge and then, by an invisible hand,

–25–

channel scarce resources to those uses socially most valuable. Private property rights have been key to economic progress worldwide since medieval times. But that conclusion needs to be confined to ordinary property rights. The latter's unambiguous desirability, while not exactly turned on its head, becomes nuanced and much less compelling when it is *intellectual* property rights that are applied to digital goods.

B Institutions and incentives

These tensions just described, between individual reward and social dissemination efficiency, are among the central forces shaping the role of digital goods in the New Economy. As societies attempt to respond to such tensions, adaptive responses emerge in the form of alternative institutions and mechanisms. Obviously, many other forces matter as well in shaping these social institutions—the goal here is only to draw one line of reasoning, not to suggest it is exclusively the sole explanation for them.

   Property rights comprise, of course, one social institution, already discussed at length in Section III.A. As described there, an explicit aim of that institution was not only to instantiate ideas, but instead to transfer and disclose them—importing for domestic use leading-edge technology and skilled artisans from abroad where they had already proven successful. Indeed, the historical evidence described by Paul David shows how even in places such as Renaissance Italy, itself a center of intellectual creativity, IPRs primarily served to encourage local introduction of innovations already made elsewhere. Thus, then as now, the element of protection in exclusivity was unquestionably present in IPRs: Individuals had to be incentivized to bring forth economically valuable ideas. However, back then but unlike now, IPRs were used to aid the spread of ideas from technologically advanced economies to those less advanced, and paradoxically to protect neither technological leadership nor the first instantiation of new ideas.

   That historical perspective has led Paul David to note how, for science and technology, two leading traditional alternatives to the institution of *property* have been *procurement* and *patronage* (hence

Paul David's taxonomy of such institutions into what he calls 3P's). These alternatives derive from viewing ideas and knowledge as classical public goods, and thus draw motivation and design from societal organizations set up for the latter.

Under procurement, a government or other wealthy benefactor sets out a specific problem that innovation should target, determines a pool of potential innovators willing to conduct the research for that problem, and then pays out of general tax revenue or benefactor wealth a designated innovator, selected from that pool of potential innovators. The chosen innovator undertakes the research and delivers the digital good for public consumption. Space and military research are examples of this. Even if military output is kept secret, it is, nonetheless, in national defense made available for public consumption.

By contrast, under patronage, a government or other wealthy benefactor establishes research awards, with parameters and goals left incompletely specified (e.g., "contribute significantly to fundamental knowledge in biology"), and allows peer-evaluated best innovators to conduct that leading-edge research with goals shared by the community of experts. Institutions such as the National Science Foundation in the US or the Economic and Social Research Council in the UK provide obvious examples of patronage. Private production of knowledge is publicly subsidized but without explicitly distorting the direction of innovation.

It is this last feature that distinguishes procurement and patronage. The former explicitly specifies the digital goods to innovate; the latter does not.

Both procurement and patronage divorce the ex ante incentive of an innovator from the ex post stream of rents generated by the innovation. Both procurement and patronage deny innovators exclusive rights to their innovations. Instead, both allow society unrestrictedly beneficial use of the innovation. The digital good is publicly disclosed: As a result, private markets in that good are shut down and the digital good disseminated as widely as ex post socially efficient. This disconnect between ex ante incentive and ex post rent has the virtue of allowing the first to remain positive and high, even as the

second converges towards zero.

The difficulty, however, is that no guidance is then provided on the true economic value of the digital good, and thus no market reveals what award, prize, or subsidy is the appropriate ex ante incentive. Admittedly, under the institution of intellectual property, that true economic value is not revealed either because the ensuing market structure is monopolistic. Uncertainty—that the digital goods uncovered by research and innovation are never exactly knowable in advance—further complicates the comparative virtues and costs of all three institutions.

Moreover, while national funding agencies—the National Science Foundation in the US or various Research Councils in the UK—can have their roles clearly explained and motivated to a tax-paying public, it is less evident that a national videogames agency or a computer software planning board would be as easily accepted. But, having said that, history provides many notable instances of patronage and procurement outside science and technology. In 1708 Johan Sebastian Bach became organist and chamber musician to the Duke of Saxe-Weimar. Apart from performing for the Duke, Bach composed numerous musical works in that employment. After 1717 when Prince Leopold appointed Bach Kapellmeister at Cöthen, Bach's duties comprised primarily instrumental composition. This royal patronage neatly divorced the pecuniary returns on Bach's music from its incentivization. In the late 18th century, wealthy patrons and the Prince-Archbishop in Salzburg and Vienna employed Mozart to produce music, to be performed subsequently in private or public. Although Mozart is reputed to have been always poor, records indicate that his income under patronage was relatively high and it was bad financial management more than insufficient compensation to blame for his poverty.

Historical reality merges features of all three different institutions for producing digital goods, and it is not always easy to describe examples as clearcut as those just given. Many research grants reward an individual academic's own research program, but just as many delineate specific goals and research questions. Universities provide a patronage umbrella for academic research, where an inno-

vator's rewards are related, loosely, to research output but where the consumers of that output do not directly pay the innovator. Some university-based innovators have taken a further step and successfully marketed their research output through private startup companies, although, varying across different countries, university administrations have progressively attempted to control that activity, shifting entrepreneurial ownership away from individual researchers and towards the university.

Adopting the Arrow-Debreu competitive equilibrium perspective in this Article makes plain that the reasoning throughout is based on economics alone: The analysis considers how alternative institutions for producing and distributing digital goods do or do not deliver socially optimal outcomes. This investigation does not take a stance on whether, say, IPRs should be awarded as a fundamental obligation or a moral right to a creator of digital goods. Social efficiency and moral rights have little to do with each other, even if these distinct dimensions to digital goods are often conflated in public discussion. Perhaps nowhere is this clearer than in the debate on Open Source Software.

C   Computer software and Open Source

Computer software is the quintessential digital good. The analysis of Section II. predicts an uneasy tension that should arise for proprietary or commercial software. This tension surfaced early in the history of commercial software when Bill Gates's 1976 open letter to computer hobbyists confronted the then-widespread practice of sharing computer code with his and Paul Allen's wanting to sell their `BASIC` interpreter for personal computers at positive price, i.e., at higher than marginal cost.

Section III.B described some societal institutions that arise in response to the difficulty in first rewarding innovators of and second distributing digital goods in general. This section argues that the Open Source Software movement arose similarly as an emergent social adaptation to the same tension between ex ante incentive and ex post

efficiency in computer software in particular: The evident success of Open Source Software draws on special characteristics of digital goods described earlier in Section II.

The organizing principles of Open Source Software center on developing and distributing computer software in specific ways. Different variants of Open Source licenses exist but all need to be approved by the Open Source Software consortium.

(Because this Article needs to focus on those issues that it has identified to be its principal concerns, it cannot do justice to many other interesting aspects to the Open Source movement, among them the history of Open Source; ongoing controversy and disagreement within and outside the movement; the movement's divergence from Richard Stallman's GNU Project; partisan internal conflicts in the use of individual charisma, community acclaim, and commercial collaboration; and varying patterns of explicit and implicit organization across 'different subgroups.)

A central requirement for Open Source Software is that the high-level language code be made available for anyone to read, modify, improve, and re-use. By contrast, almost all commercial software is distributed only as a machine-readable binary executable, with severe licensing restrictions on how it may be used. Commercial software, in other words, is a digital good where intellectual property rights—usually copyright, but increasingly patent as well—hide and protect the underlying bitstring.

(The concept of a digital good used in this article—namely, a string of 1s and 0s—needs refinement when analyzing Open Source Software. Both high-level language source and machine binary versions of computer software are, in our definition, digital goods. But the former is structured English prose, which is rendered into 1s and 0s for storage in a computer, and can be transparently moved across systems and hardware. The latter, by contrast, exists *only* as a string of 1s and 0s, and is specific to particular computer processors and hardware. High-level language source code is how programmers write software; that source code can be read and modified readily by any human with basic computer skills. Humans can neither read nor write machine binary executables or binaries, outside the most trivial

instances. Machine binaries result from compiling and linking source code on specific computer hardware.)

Development of Open Source Software typically begins with an announcement on Internet news groups of either availability or intent of computer software to achieve a particular purpose. Interested participants worldwide express interest and offer assistance in the form of computer code already written; or download, when available, extant source code or binary executables for certain common computer platforms (e.g., Intel machines running the `GNU/Linux` operating system [OS]). Users push the software to its limits and uncover deficiencies, that can then be either reported back to the central coordinators— again by Internet news groups or by email—or fixed by the users themselves. Either way, if the fixes (or patches) are judged appropriate when communicated back to a coordinator, they are folded into the ever-changing distribution source tree. The process continues with no natural endpoint; hundreds, potentially thousands, of programmers, at varying levels of expertise, participate.

Every step of this aspatial but worldwide development is undertaken electronically over the Internet. In many instances, there is minimal human intervention, with computer software delivering and seamlessly merging patches into the underlying source tree for redistribution. Open Source Software is, therefore, distributed freely—i.e., at price equal only to the cost of download, installation, and familiarity (all of which arise as well with commercial software, on top of the latter's sticker price)—and with much less severe licensing restrictions than commercial software. Participants contributing code in Open Source Software receive no direct pecuniary compensation tied to the distribution of the digital good that is the software product.

Observers have remarked how rapidly a robust software product emerges from this decentralized and only lightly-coordinated process, very different from traditional modes of software development. Notable examples of Open Source Software projects that have, by some measures, out-performed their commercial counterparts include `GNU/Linux` (a `Unix` OS together with a massive complement of software tools, judged by many to be the only serious competitor to Microsoft's Windows OS); the `Apache` webserver (over half the In-

ternet); and `sendmail` (the single most widely-used email transport agent). The Internet Operating System Counter[1] found `GNU/Linux` accounted for 31.3% of 1.25 million Internet-connected hosts recognized (out of 1.5 million processed) in April 1999, an increase from 28.5% over the corresponding ratio in January 1999. By contrast, Microsoft Windows hosts declined from 24.4% to 24.3% over this same period. In that period `GNU/Linux` accounted for both the single most common host OS and the fastest growth rate across host operating systems. Previously purely commercial ICT companies have adopted instances of Open Source Software licensing: High-profile examples include Sun Microsystems, Nokia, Intel, and IBM.

Two questions matter importantly in the economics here. First, what incentives and constraints inform the decisions made by individual programmers and companies who produce code for Open Source Software? For the most part these agents appear to work for no obvious direct compensation, but instead choose to supply computer code freely for unfettered redistribution to the worldwide community. This question has been considered by economists such as Josh Lerner and Jean Tirole, and by Open Source Software participants themselves: One answer is that compensation is a dynamic process so that individual acclaim and reputation-building—successfully submitting code that meets the rigorous standards of excellence demanded by the Open Source community—increases the individual's likelihood of accession to high-paying software employment. In this view, current measured compensation does not meaningfully measure economic reward as longer-term career concerns motivate the individual. This temporal disconnect between effort and reward explains why workers only appear to contribute time and code for free.

The second question is much less studied by economists but perhaps more intriguing: Why does such a system of decentralized, uncoordinated software production and distribution succeed—and, by some measures, succeed so spectacularly? Leave aside where individual motivation comes from; why does apparently haphazard Open

---

[1] `http://www.leb.net/hzo/ioscount/`, accessed 23 September 2002

Source collaboration produce good software? Understanding the microeconomics in the motivation of individual programmers is an important first step in understanding Open Source. But then, taking the incentive as given, why does the process succeed? After all, if reputation and long-term career concerns only substitute for the more transparent and typical direct compensation, then Open Source Software should be consistently neither more nor less successful than commercial software operations. Indeed, insofar as these less-transparent nonpecuniary mechanisms engender greater transaction and monitoring costs than simple, ordinary pecuniary exchange, even if both Open Source and commercial software are observed in practice, the former should manifest only transitorily before more economically efficient commercial operations again dominate in the long run.

To address this second question, some observers have attempted to draw a connection between software and the theory of emergent evolution in complex adaptive systems. The key idea is that for complex adaptive systems a surprising global property can emerge out of uncoordinated, individual local actions—even, or especially, without centralized coordination. Thus, large complicated software systems like `GNU/Linux`, `Apache`, and others emerge globally robust simply from many, many programmers individually working away on small, isolated facets of a large project. The system is complex and adaptive.

Here, I consider these developments in relation to the discussion on the economics of digital goods developed earlier in this article. What is distinctive about the voluntary collaborative production of software that differentiates it from other voluntary collaborations such as, say, working in the Peace Corps (where, to some degree, reputation and long-term career concerns likely figure as well)?

To begin, recall that the Fundamental Theorems of Welfare Economics already provide a statement similar to that for emergence in complex adaptive systems: Global efficiency arises, seemingly spontaneously, from self-seeking, uncoordinated, decentralized, local price-taking behavior. Thus, the Fundamental Theorems of Welfare Economics apparently predict how, in Eric Raymond's evocative language, uncoordinated actions in the apparently haphazard Bazaar produce that outcome that the centralized Cathedral seeks—there is

no contradiction between the Cathedral and the Bazaar.

Problem is, the market for the digital good that is computer software violates the usual conditions for applying these Theorems. As previously discussed, intellectual property rights over digital goods differ from regular property rights over ordinary goods in the Arrow-Debreu model. Exchange of digital goods is unlike that of ordinary goods. Monopolistic licensing restricts trade and therefore produces allocations sharply different from those under perfect competition. Thus, even if individual incentives suffice to drive companies and workers to code Open Source Software projects, that those incentives exist cannot, by itself, explain why Open Source products are successful relative to other models of production.

Yet those same properties of digital goods that render invalid the standard Welfare Theorems might also, at the same time, be what make possible the Open Source movement's success. Three distinct dimensions underly this conjecture. First, the Open Source movement is, in the main, an attempt to circumvent traditional IPR features in commercial software. One of the movement's organizing principles targets specifically distribution mechanisms that reinstate the ex post social efficiency condition that marginal cost equals marginal benefit. It thus, consciously or otherwise, attempts to resurrect the Fundamental Theorems of Welfare Economics.

Second, observers have noted that uncovering software bugs and studying, testing, and experimenting with Open Source code is a worldwide process involving thousands of participants—even if, for certain specific Open Source projects, code that ultimately survives into the final distribution tree can be directly traced to only a very few extraordinarily-gifted lead programmers. [In the words of Eric Raymond, "Given enough eyeballs, all bugs are shallow (Linus's Law)".] Drawing on the very large base of contributors is enabled by aspatial atemporal collaboration over the Internet. Such collaboration would be infeasible if the output being studied and tested were not nonrival and easily and (for all practical purposes) instantaneously transportable globally. Put more positively, because software is aspatial and nonrival, its use and testing can proceed costlessly in parallel. Improvement through intensive use is therefore rapid to a degree un-

available with nondigital goods.

Third, this rapid improvement is facilitated by yet another digital good property, namely, that the product in source code is a human-readable recipe—understandable and malleable—allowing recombinant development. Alternative experimental perturbations can be quickly and costlessly tested. Although the space of possible variations in the software is extremely large, simultaneous parallel processing by as large a user base as possible allows rapidly locating productive and successful directions for further development. If we describe the use of software as a set of three bitstrings, the program itself being the first, input actions the second, and finally output actions the third, clearly the range of possible outputs is made larger (usually by orders of magnitude) when the program bitstring can be altered simultaneously with the set of input bitstrings. This variation thus further enlarges recombinant possibilities. By contrast, keeping the code confidential and disseminating to users only a closed black box, as happens with commercial software, removes the possibility of ongoing recombinant development.

To be clear, this Article does not pretend to suggest that the economics of digital goods alone completely explains the success of Open Source Software. Rather, it has sought to describe where that economics helps and where it doesn't. That individual incentives in the Open Source movement might exist can be only part of the explanation. Needed on top of that still is a complete economic equilibrium description of how successful outcomes can emerge from the complex range of individual actions and mass interactions undertaken in Open Source Software.

D   Geography

The rising importance of digital goods in the New Economy further eases how ideas and goods can be transported across space. In the perspective developed in this Article, it is not just that transportation technology is improving, given an invariant set of objects to be transported. Instead, that to be transported that is economically

valuable is itself evolving, towards ever greater geographical mobility. In the limit, digital goods are aspatial—they are at once everywhere and nowhere.

But this hypothesized aspatial character to digital goods—an analytical description based on their physical (or, perhaps more correctly, nonphysical) nature—seems to collide dramatically with empirical observation. What both casual empiricism and rigorous empirical analysis suggest is not that space doesn't matter but, instead, the opposite: Spatial concentration is the single most distinctive feature of all economic activity, including notably that activity that is knowledge-intensive. Indeed, the geographical clustering of computer software and digital media production, academic and commercial R&D, and financial services, among other digital goods, is likely tighter than for ordinary goods and services. Does this mean that knowledge spillover is only geographically localized, that digital goods have restricted spatial reach and thus are not aspatial after all?

To resolve this apparent paradox, notice that aspatiality in digital goods does not imply space no longer matters. Instead, the correct inference is only that for digital goods transportation costs don't matter, so that all *other* reasons for why geography is important now assume heightened significance. Thus, it does not refute aspatiality in digital goods to observe that particular items of scientific knowledge are most intensely shared among researchers in a relatively small geographical area (Silicon Valley; Washington DC and northern Virginia; Cambridge England; Route 128 Massachusetts; Bangalore India), that creative media industries are spatially clustered (Soho in New York City; Shoreditch and Islington in London), or that computer software can be used only with computer hardware that must sit *somewhere*.

These examples and similar others show that digital goods are often used or consumed using complementary inputs. For scientific knowledge, researchers whose embodied human capital apply knowledge to its most productive use cannot locate nowhere. Those researchers must work somewhere and so might well cluster geographically because communication of tacit knowledge, not digital goods, is most efficient in close physical proximity. For similar reasons, creative

artists might locate jointly in spatial proximity. Neither example refutes that codified knowledge, the digital good, is aspatial. Indeed, it is that aspatiality that elevates the importance of all the other reasons for complementary inputs to co-locate and that thus, seemingly paradoxically, induces spatial clustering in the use of aspatial digital goods.

Consider the resources expended in transporting ordinary heavy output to consumers from where that output is produced: If such costs explain why production spreads out across geography—because production needs to move partway towards where consumers live— the increasing importance of digital goods implies higher spatial clustering, not greater dispersion. If synchronous face-to-face interactions matter for transmitting tacit (nonbitstring) knowledge, then the growing significance of aspatial digital goods raises the rent on such tacit knowledge, increasing the importance of localized face-to-face communications and thereby raising spatial concentration.

A more cogent objection to aspatiality is to note that surrounding the production and distribution of digital goods are technologies to improve data compression and increase transmission bandwidth. These auxiliary technologies are needed to transport digital goods— they would be unnecessary if digital goods were aspatial. But the effect of these technologies is precisely to bring about that aspatiality. That these technologies succeed means that consumers and producers (not themselves in such industries) can treat digital goods as effectively aspatial.

When Alfred Marshall described how industrial centers ferment ideas so that "mysteries of the trade become no mystery; but are as it were in the air", and how "if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas", Marshall was attempting to make sense of the particular industrial clustering he observed around him. However, the notion of "ideas in the air" seems to have taken a varied life of its own, and is routinely used to deny that ideas and knowledge might be aspatial and to assert instead that they must be geographically localized.

But Marshall's vivid phrasing, removed from the very specific

particulars of late 19th-century England, can be just as forcefully deployed to visualize how Kurt Gödel in Vienna, Alan Turing in Cambridge England, Emil Post in New York, and Alonzo Church in Princeton—geographically far apart but almost simultaneously in time—developed closely-related tools and results in mathematical logic in the mid 1930s. Such "Merton multiples" abound in the history of scientific and artistic discovery: They are distinguished by being close together in time, not in space.

To summarize, completely prosaic reasons, simultaneously with aspatiality in digital goods, might underly why spatial clustering occurs even in knowledge-intensive industries that either take aspatial digital goods as inputs or produce them as outputs. Precisely because these inputs or outputs are aspatial, their location-determining role vanishes from the equation, and it is the other, more straightforward and more traditional forces that assume greater significance. Spatial clustering in knowledge-intensive activity can, therefore, provide evidence showing how digital goods are importantly aspatial and not the opposite, that digital goods have only restricted geographical reach.

This reasoning has one obvious counter-example of note. What if there is no traditional factor input, so that it is only aspatial digital goods interacting with yet other aspatial digital goods in production and consumption? Can clustering then spontaneously emerge? In other words, can distinct geographical patterns that are not trivial— all activity concentrated in just one place or, alternatively, all activity randomly distributed (spatial white noise)—then arise? If so, what determines those emergent spatial distributions?

One possible answer is that localization in time, in the sense of global timezones (not calendar time), rather than geographical distance might then become the critical spatial feature. The current author has modelled analytically such possibilities, using ideas from modern economic geography and Turing's theory of morphogenesis. This reasoning suggests a more apposite test to assess if digital goods are aspatial. Economic activity purely in digital goods—where transportation costs do not matter but communication synchronicity does—should have global clustering line up longitudinally, rather than along the two dimensions running across the Earth's surface.

Spatial clustering should be latitude-blind. Or, put yet differently, distance in space is irrelevant but separation across timezones matters. Certainly, editorial staff for newspapers with global distribution now organize along timezones, and traders in financial markets keep an eye on timezones, not on latitude. Intel's integrated circuit design facilities spanning Israel and California are organized with an eye to passing on work across timezones. Bangalore is able to provide overnight medical transcription service for physicians in the US because of its longitudinal location, not its latitudinal one.

IV.   Conclusions

This Article has considered digital goods in the New Economy, and described several economic implications of their growing importance. It has provided some scientific, social, and historical background to this ongoing evolution.

Instead of hypothesizing ad hoc implicit economic frictions that the New Economy can then purport to overcome, this Article instead adopted a base perspective of markets in perfectly competitive equilibrium, and asked, What is distinctive about the New Economy in general or digital goods in particular that might affect economic performance?

The Article has described how digital goods are nonrival, infinitely expansible, discrete, aspatial, and recombinant. Often a number of these properties are simply grouped together interchangeably under the term "increasing returns". This Article has attempted to show how doing so can mislead.

This Article has attempted to describe how, given our current state of knowledge, digital goods and the New Economy make for implications truly different from what we have traditionally understood. A compact summary of the principal conclusions might be useful here: Intellectual property rights have far from the same compelling justification that ordinary property rights do, in providing for incentive and efficiency in economic systems. Historically, different institutions have emerged to circumvent social inefficiencies due to the peculiar

properties of intellectual assets. Those same peculiar properties might currently be inducing new institutions, such as the Open Source Software movement, with features especially and spontaneously tuned to deal with digital goods. Finally, disappearing transportation costs will affect the location of high-value economic activity in geography and space, although not always in the most obvious ways.

Many issues remain to studied rigorously, and many intellectual connections to be discovered. Until these complications are better understood, working out the implications of digital goods in the New Economy for productivity and growth or competitiveness and social equity will likely be delicate.

Bibliography

Arrow, Kenneth J. (1962) "Economic welfare and the allocation of resources for inventions," In *The Rate and Direction of Inventive Activity,* ed. Richard R. Nelson (Princeton: Princeton University Press and NBER) pp. 609–625

Arthur, W. Brian (1994) *Increasing Returns and Path Dependence in the Economy* (Ann Arbor: The University of Michigan Press)

Boldrin, Michele, and David K. Levine (2002) "The case against intellectual property," *American Economic Review (Papers and Proceedings)* 92(2), 209–212, May

Coase, Ronald W. (1960) "The problem of social cost," *Journal of Law and Economics* 3, 1–44, October

Cowan, Robin, Paul A. David, and Dominique Foray (2000) "The explicit economics of knowledge codification and tacitness," *Industrial and Corporate Change* 9(2), 211–253, June

Dasgupta, Partha (1988) "Patents, priority and imitation or, the economics of races and waiting games," *Economic Journal* 98, 66–80, March

Dasgupta, Partha, and Eric Maskin (1987) "The simple economics of research portfolios," *Economic Journal* 97, 581–595, September

David, Paul A. (1992) "Knowledge, property, and the system dynamics of technological change," *Proceedings of the World Bank Annual Conference on Development Economics* pp. 215–248, March

——— (1993) "Intellectual property institutions and the panda's thumb: Patents, copyrights, and trade secrets in economic theory and history," In *Global Dimensions of Intellectual Property Rights in Science and Technology,* ed. M. B. Wallerstein, M. E. Mogee, and R. A. Schoen (Washington DC: National Academy Press) chapter 2, pp. 19–61

Enderton, Herbert B. (1972) *A Mathematical Introduction to Logic* (New York: Academic Press)

Evans, David S. (2001) "Is free software the wave of the future?," *The Milken Insitute Review* 3(4), 34–41, Fourth Quarter

Fujita, Masahisa, Paul Krugman, and Anthony Venables (1999) *The Spatial Economy: Cities, Regions, and International Trade* (Cambridge: MIT Press)

Glaeser, Edward L., Hedi D. Kallal, José A. Scheinkman, and Andrei Shleifer (1992) "Growth in cities," *Journal of Political Economy* 100(4), 1126–1152, December

Helpman, Elhanan, ed. (1998) *General Purpose Technologies and Economic Growth* (Cambridge: MIT Press)

Hirshleifer, Jack, and John G. Riley (1992) *The Analytics of Information and Uncertainty* Cambridge Surveys of Economic Literature (Cambridge England: Cambridge University Press)

Hofstadter, Douglas R. (1979) *Gödel, Escher, Bach* (New York: Basic Books)

Holland, John H. (1995) *Hidden Order: How Adaptation Builds Complexity* (Reading Massachusetts: Addison-Wesley)

Holyoak, Jon, and Paul Torremans (1995) *Intellectual Property Law* (London: Butterworths)

Jaffee, Adam B., Manuel Trajtenberg, and Rebecca Henderson (1993) "Geographic localization of knowledge spillovers as evidenced in patent citations," *Quarterly Journal of Economics* 108(3), 577–598, August

Jones, Charles I. (2002) *Introduction to Economic Growth,* second ed. (New York: W. W. Norton)

Kauffman, Stuart A. (1993) *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford University Press)

Kolko, Jed (2002) "Silicon mountains, silicon molehills: Geographic concentration and convergence of Internet industries in the US," *Information Economics and Policy* 14(2), 211–232, June

Kretschmer, Martin, George Michael Klimis, and Roger Wallis (2001) "Music in electronic markets," *New Media and Society* 3(4), 417–441

Krugman, Paul (1991) *Geography and Trade* (Cambridge: MIT Press)

Lerner, Josh, and Jean Tirole (2002) "Some simple economics of Open Source," *Journal of Industrial Economics* 50(2), 197–234, June

Lipscomb, Andrew A., and Albert Ellery Bergh, eds (1905) *The Writings of Thomas Jefferson* Thomas Jefferson Memorial Association (Chicago: The University of Chicago Press)

Moody, Glyn (2001) *Rebel Code: Linux and the Open Source Revolution* (London: Penguin)

Nagel, Ernest, and James R. Newman (1958) *Gödel's Proof* (New York: New York University Press)

Nordhaus, William D. (1969) *Invention, Growth, and Welfare: A Theoretical Treatment of Technological Change* (Cambridge: MIT Press)

North, Douglass C. (1981) *Structure and Change in Economic History* (New York: Norton)

Quah, Danny (2000) "Internet cluster emergence," *European Economic Review* 44(4–6), 1032–1044, May

—————— (2001) "The weightless economy in economic development," In *Information Technology, Productivity, and Economic Growth,* ed. Matti Pohjola UNU/WIDER and Sitra (Oxford: Oxford University Press) chapter 4, pp. 72–96

—————— (2002a) "24/7 competitive innovation," Working Paper, Economics Dept., LSE, London, May

—————— (2002b) "Almost efficient innovation by pricing intellectual assets," Working Paper, Economics Dept., LSE, London, June

—————— (2002c) "Matching demand and supply in a weightless economy: Market-driven creativity with and without IPRs," *De Economist* 150(4), 381–403, October

—————— (2002d) "Spatial agglomeration dynamics," *American Economic Review (Papers and Proceedings)* 92(2), 247–252, May

—————— (2002e) "Technology dissemination and economic growth: Some lessons for the New Economy," In *Technology and the New Economy,* ed. Chong-En Bai and Chi-Wa Yuen (Cambridge: MIT Press) chapter 3, pp. 95–156

Raymond, Eric S. (2000) "The cathedral and the bazaar," Webpages, `http://www.tuxedo.org/~esr/writings/cathedral-bazaar/`. accessed: 11 November 2002

Romer, Paul M. (1990) "Endogenous technological change," *Journal of Political Economy* 98(5, part 2), S71–S102, October

—————— (1994) "New goods, old theory, and the welfare costs of trade restrictions," *Journal of Development Economics* 43(1), 5–38

Rosen, Sherwin (1981) "The economics of superstars," *American Economic Review* 71(5), 845–858, December

Saxenian, Annalee (1994) *Regional Advantage: Culture and Competition in Silicon Valley and Route 128* (Cambridge: Harvard University Press)

Scotchmer, Suzanne (1995) "Patents as an incentive system," In *Economics in a Changing World,* ed. Jean-Paul Fitoussi, vol. 5 of *Proceedings of the Tenth World Congress of the International Economic Association, Moscow* (London: St Martin's Press) chapter 12, pp. 281–296

Shapiro, Carl, and Hal R. Varian (1999) *Information Rules: A Strategic Guide to the Network Economy* (Boston: Harvard Business School Press)

Turing, Alan M. (1952) "The chemical basis of morphogenesis," *Philosophical Transactions of the Royal Society of London Series B* 237, 37–72, August

Weitzman, Martin L. (1998) "Recombinant growth," *Quarterly Journal of Economics* 113(2), 331–360, May

Wright, Brian D. (1983) "The economics of invention incentives: Patents, prizes, and research contracts," *American Economic Review* 73(4), 691–707, September

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

548	T. Bayoumi M. Haacker	It's Not What You Make, It's How You Use IT: Measuring the Welfare Benefits of the IT Revolution Across Countries
547	A. B. Bernard S. Redding P. K. Schott H. Simpson	Factor Price Equalization in the UK?
546	M. Guti鐁rrez-Dom鐁nech	Employment Penalty After Motherhood in Spain
545	S. Nickell S. Redding J. Swaffield	Educational Attainment, Labour Market Institutions and the Structure of Production
544	S. Machin A. Manning J. Swaffield	Where the Minimum Wage Bites Hard: the Introduction of the UK National Minimum Wage to a Low Wage Sector
543	R. Belfield D. Marsden	Matchmaking: the Influence of Monitoring Environments on the Effectiveness of Performance Pay Systems
542	C. A. Pissarides	Consumption and Savings With Unemployment Risk: Implications for Optimal Employment Contracts
541	M. Amiti C. A. Pissarides	Trade and Industrial Location with Heterogeneous Labor
540	G. Duranton H. G. Overman	Testing for Localisation Using Micro-Geographic Data
539	D. Metcalf	Unions and Productivity, Financial Performance and Investment: International Evidence
538	F. Collard R. Fonseca R. Mu鐁oz	Spanish Unemployment Persistence and the Ladder Effect
537	C. L. Mann E. E. Meade	Home Bias, Transactions Costs, and Prospects for the Euro: A More Detailed Analysis
536	M. Manacorda E. Moretti	Intergenerational Transfers and Household Structure. Why Do Most Italian Youths Live With Their Parents?