# 5

# Nontraditional Infrastructure

IN THE NEXT TWO CHAPTERS, I shift attention from traditional infrastructure to nontraditional infrastructure, specifically, from transportation and communications infrastructure to environmental and intellectual infrastructure. It may seem odd to be grouping roads and telephone networks with lakes and ideas under the infrastructure umbrella. The primary reason for doing so is to highlight the demand-side similarities and the important, if varied, role of commons management. When feasible, society benefits tremendously by leveraging nonrivalry to support nondiscriminatory access to such resources, because doing so enables the public to participate productively in a wide range of socially valuable activities. As with traditional infrastructure, many environmental and intellectual infrastructure resources are public, social, and mixed infrastructures that contribute immensely to our economic and social development. In fact, there are interesting parallels between environmental and intellectual infrastructure resources: Both are inputs into complex dynamic processes—natural ecosystem processes and cumulative intellectual processes, as well as social and cultural processes—that have the potential to yield significant positive externalities that benefit society as a whole. Sustaining these fundamental resources in an open manner is critical to realizing this potential.

Feasible leveraging depends, however, on managing a host of competing considerations. Intellectual infrastructures face supply-side issues similar to those issues faced by traditional infrastructure. Attracting private investment can be difficult because of the cost structure of supply, high costs of exclusion, and misappropriation risks, among other factors discussed below. Environmental infrastructures do not face the same supply-side issues. Human beings do not supply environmental infrastructure in the same manner as traditional infrastructure or intellectual infrastructure. We inherit and manage environmental resources.[1] But, like traditional infrastructure, environmental infrastructures are partially (non)rivalrous and, as a result, face complex congestion and degradation problems. In short, pure open access to intellectual or environmental infrastructure typically is not feasible absent additional institutional support, whether in the form of public subsidies for basic research or in the form of command-and-control regulation of industrial polluters.

Viewing environmental and intellectual infrastructure through the infrastructure lens yields important insights regarding commons management institutions. In particular, both environmental and intellectual property legal systems construct semi-commons arrangements that create and regulate interdependent private rights and public commons. Each system does so in a very different way, as explored below. Though these areas of law are often conceptualized as necessary departures from commons because of the tragic effects that commons might have on incentives—whether leading to overconsumption of environmental resources or undersupply of intellectual resources—greater analytical clarity may be achieved by conceptualizing these legal regimes in a more nuanced fashion and appreciating that an overarching objective and consequence of these regimes is sustaining (rather than eliminating) commons for an appreciable range of use(r)s.

[1] Of course, each generation also inherits an incredible wealth of intellectual and cultural resources—our cultural environment—and as with the inherited natural environment, these inherited resources are the foundation for our current reality and future development and progress as a society.

# 12

# INTELLECTUAL INFRASTRUCTURE

THIS CHAPTER EXPLORES how infrastructure theory applies to cultural-intellectual resources and delineates a class of infrastructure—hereinafter, *intellectual infrastructure*. Intellectual infrastructure, such as basic research, ideas, general purpose technologies, and languages, creates benefits for society primarily by facilitating a wide range of downstream productive activities, including information production, innovation, and the development of products and services, as well as education, community building and interaction, democratic participation, socialization, and many other socially valuable activities. The foundational role of intellectual infrastructures in cumulative, dynamic, and complex systems merits attention. Courts and commentators frequently refer to intellectual infrastructure resources as "building blocks" to capture their role as basic inputs. But while the "building blocks" metaphor is evocative, it fails to fully reflect the complex relationships among participants in intellectual systems that derive value from intellectual infrastructures as producers, users, consumers, or incidental beneficiaries.

Applying infrastructure concepts to cultural-intellectual resources is more difficult in some respects than applying them to other resources. The difficulties stem in part from the fluid, continuous, and dynamic nature of cultural-intellectual systems. Distorting reductionism seems inevitable when we attempt to delineate clear boundaries around discrete cultural-intellectual resources or to separate resources (inputs, outputs, products, things) from activities (processes, practices, uses). Intellectual infrastructures, such as

basic research, often seem to be both resources and activities. Difficulties also stem from the fact that infrastructure appear to exist on many different scales within cultural-intellectual systems. Nonetheless, the analysis yields important insights about societal demand for intellectual infrastructure and the case for commons management.

This chapter is organized into four sections. It begins in section A with the idea of the cultural environment as infrastructure. This discussion provides an important connection to the previous chapter and establishes a context for examining intellectual-cultural resource systems and governance institutions. It also explains some of the complex relationships between intellectual-cultural resources and people—for example, the mutual dynamic shaping that takes place as society lives within, interacts with, shapes, and is shaped by its interactions with the cultural environment. Section B describes the economic characteristics of intellectual resources. The economic analysis of intellectual resources is quite complex; while I discussed most of the general issues in previous chapters, this section extends the analysis in a few ways. First, it explains how (non) excludability and nonrivalry[1] give rise to distinct economic considerations concerning systematic risk of undersupply of some intellectual resources. Next, it considers an added layer of complexity associated with "information [being] both input and output of its own production process. . . . This characteristic is known to economists as the 'on the shoulders of giants' effect."[2] The added complexity centers on the dynamics of intellectual processes and systems and how intellectual progress occurs. An appreciation of the "on the shoulders of giants" effect is critical to understanding how productive use of intellectual resources generates spillovers. It also reconnects the economic analysis with the broader notion of the cultural environment.

[1] I use (non)excludability because this characteristic is context-dependent, is variable with the costs of exclusion, and can be addressed though various institutional interventions. I use nonrivalry (without the parentheses) because this characteristic in inherent or fixed for intellectual resources. While some dispute this point and argue that I am too much of a Platonist, I have never found this argument persuasive. The marginal cost of allowing another person to consume an intellectual resource is always zero because the resource has infinite capacity to be possessed, consumed, and used; the resource is not depleted. This does not mean that each possessor, consumer, or user will realize the same (or even any) value. An idea might bring me some positive value (benefit), someone else no value, and someone else negative value (harm). Regardless, the idea is nonrival and sharable. If I write a PhD thesis and capture all of the value from an idea that can be realized in the process of being awarded a PhD and I effectively preclude someone else from using the idea for her PhD, the preclusive effect is a function of the external environment, the PhD rules or market, or whatever one calls it, but this does not mean the idea is anything other than nonrival. A second comer could possess the idea, and even write a PhD thesis, but the person would not, according to the rules, be awarded a PhD. This may reduce the use value realized by a second user who wishes to make a particular use of the idea, but that has nothing to do with nonrivalry. The PhD queue may be congestible, but the idea is not. Similarly, a person might need a certain level of knowledge or language skills (absorptive capacity) to effectively use an idea, but that important prerequisite also has nothing to do with nonrivalry.

[2] Benkler 37 (2006a).

I argue that the complexity of dynamic cultural systems and the resulting economics should lead to a reframing of the economic objectives in this area. Specifically, I suggest that pursuing optimal production of public goods is a fool's errand because of the pervasiveness and diversity of externalities in the cultural environment. It is more appropriate to improve the efficiency of inframarginal investments,[3] for example, by reducing the costs of public goods production for as wide a variety of potential public goods producers as possible. This shift in perspective has implications for the subsequent discussion of intellectual infrastructure (section C) and intellectual property laws as semi-commons arrangements (section D).

Section C focuses on applying the infrastructure criteria to delineate intellectual infrastructure. Through a series of examples, this section describes some of the practical difficulties to doing so. The section then turns to ideas as an example of intellectual infrastructure. Focusing on the First Amendment, copyright law, and patent law, it examines legal recognition of both the infrastructural nature of ideas and the social value of commons management. It then considers the struggle in patent law to differentiate abstract ideas from patentable inventions, reflected most recently in the Supreme Court's decision in *Bilski v. Kappos* (2010), and suggests that patent law should follow more explicitly the analytical framework employed in copyright law. I argue that despite some confusion and controversy on how to draw lines and separate ideas from expression and invention, ideas are and should be "free as the air to common use."[4]

Section D considers intellectual property laws. It examines intellectual property laws as a semi-commons regime and compares it to the regulatory semi-commons discussed in the previous chapter. It shows first how intellectual property laws, like environmental regulation, are targeted exceptions/interventions to the default commons regime for the cultural environment. The laws enclose and regulate a select (albeit very broad) set of intellectual resources to overcome supply-side market failures. Given tremendous difficulties in establishing and maintaining boundaries around this set and the dynamic and complex nature of cultural-intellectual resource systems, the laws also sustain semi-commons arrangements for enclosed resources. This leads to a second point. Many intellectual infrastructure resources are excluded from the enclosed set—that is, the resources remain in the public domain and are not patentable or copyrightable, but many intellectual infrastructures are patentable or copyrightable. For these resources, intellectual property laws seem to recognize the social demand for commons management and mediate access to intellectual infrastructure accordingly. I argue that a more deliberate approach is needed and provide some preliminary suggestions.

[3] See chapter 3, appendix.

[4] International News Serv. v. Associated Press, 248 U.S. 215, 250 (1918) (Brandeis, J., dissenting).

## A. The Cultural Environment as Infrastructure (Meta- or Infra-infrastructure)

We can identify infrastructure at a very high level of abstraction and broad scale—the cultural environment—and at progressively lower levels of abstraction, on smaller scales, and within different systems or fields. In parallel with such efforts, we can study different commons management institutions. Thus, we analyze both infrastructure and commons as nested phenomena operating at different levels that may interact with one another.[5] The scaling issue arises in other contexts—for example, recall how our discussion of road infrastructure varied in scale from a particular road or bridge to the national highway system.[6]

At a relatively abstract level, the basic similarities between the natural and cultural environments concern the functional and relational meanings of the common term "environment." An environment might be defined as a complex system of interconnected and/or interdependent resources (or even resource systems) that comprise the "surroundings," "setting," or "context" that we inherit, live within, use, interact with, change, and pass on to future generations. We inherit the natural physical environment; we live within, use, interact with, and change it; and we pass it on to future generations. Similarly, we inherit, live within, use, interact with, change, and pass on to future generations a cultural-intellectual environment, comprised of many overlapping sub-environments, if one would like to distinguish culture(s), science(s), and so on. The world we live in comprises multiple, complex, overlapping, and interdependent resource systems with which we interact and that constitute our environments—the natural environment is one type, and (socially) constructed environments, such as the cultural environment, are another.

Thus, we can envision a cultural environment that consists of the various cultural, intellectual and social resource systems that we inherit, use, experience, interact with, change, and pass on to future generations.[7] This is analogous to the move made in the previous chapter envisioning the environment as a meta–natural environment that consists of various overlapping and interdependent natural resource ecosystems.[8]

---

[5] Madison, Frischmann, & Strandburg 674 (2010) ("By 'nesting,' we mean that a particular commons phenomenon might be analyzed at many levels; these levels may interact strongly with one another. One of the issues that must be resolved in any particular inquiry is the appropriate level of complexity at which a particular commons should be studied. Ostrom analogizes this nested analysis to a set of maps at different levels of detail, such as one sees when using the zoom function in Google Maps. All of these maps are accurate, but the usefulness of a particular map depends on the question one seeks to answer. Moreover, some questions can be answered by focusing only at street level, while others may require zooming back and forth to different levels. Similarly, analyzing a commons institution may require more or less detailed knowledge of the larger cultural institutions within which it resides.") (citing Ostrom 58–62 (2005)).

[6] See chapters 7, 8, and 9 (identifying infrastructure at different scales and levels of abstraction).

[7] I explored this is greater detail in prior work; see Frischmann 1091–96 (2007a); Madison, Frischmann, & Strandburg 685 (2010). On cultural environmentalism, see Boyle 70–74 (2003); Boyle 108–16 (1997); Boyle (1996); Opderbeck (2009).

[8] I owe a significant intellectual debt to Jamie Boyle for his work on cultural environmentalism. I discussed his work extensively in Frischmann (2007a).

The cultural environment provides us with resources and capabilities to act, participate, be productive, and "make and pursue life plans that can properly be called our own."[9] It also shapes our very beliefs and preferences regarding our lives (life plans) and relationships with each other and the world we share. Human beings are not born with fully formed preferences, knowledge, and beliefs about the world they enter;[10] rather, preferences, knowledge, and beliefs are learned and experienced and thus contingent to a degree on the cultural environment a person experiences.[11]

We have an incredibly complex and dynamic relationship with the cultural environment.[12] Science and culture, for example, are cumulative and immersive systems that develop with society, while simultaneously developing society. Put another way, the cultural environment provides for, shapes, and reflects us, and at the same time, we provide, shape, and reflect it. I stress this point because the cultural environment has a normative dimension that is sometime lost. As John Breen puts it, culture can be understood as a society's answer to a series of "fundamental questions" about what it values. He explains:

---

[9] Benkler 146 (2006a).

[10] DeMartino 77–79 (2000); Krackhardt 5 (1994); North 46–47 (2005). I have always loved how Julie Cohen made this point in an article about privacy:

> Autonomous individuals do not spring full-blown from the womb. We must learn to process information and to draw our own conclusions about the world around us. We must learn to choose, and must learn something before we can choose anything. Here, though, information theory suggests a paradox: "Autonomy" connotes an essential independence of critical faculty and an imperviousness to influence. But to the extent that information shapes behavior, autonomy is radically contingent upon environment and circumstance. The only tenable resolution—if "autonomy" is not to degenerate into the simple, stimulus-response behavior sought by direct marketers—is to underdetermine environment. Autonomy in a contingent world requires a zone of relative insulation from outside scrutiny and interference—a field of operation within which to engage in the conscious construction of self. The solution to the paradox of contingent autonomy, in other words, lies in a second paradox: To exist in fact as well as in theory, autonomy must be nurtured.
>
> A realm of autonomous, unmonitored choice, in turn, promotes a vital diversity of speech and behavior. The recognition that anonymity shelters constitutionally-protected decisions about speech, belief, and political and intellectual association—decisions that otherwise might be chilled by unpopularity or simple difference—is part of our constitutional tradition. But the benefits of informational autonomy (defined to include the condition in which no information is recorded about nonanonymous choices) extend to a much wider range of human activity and choice. We do not experiment only with beliefs and associations, but also with every other conceivable type of taste and behavior that expresses and defines self. The opportunity to experiment with preferences is a vital part of the process of learning, and learning to choose, that every individual must undergo.

Cohen (2001) (footnotes omitted). Cohen's argument strongly influenced my own thinking.

[11] Of course, not everyone experiences the same cultural environment. Individual cultural experience is inextricably linked to the extent of one's access to artistic and cultural resources; these resources are distributed spatially in ways that make any particular resource more relevant or less so to a given individual, based on accessibility and proximity. Cohen 1180 (2007).

[12] Breen 29–30 (2006) (quoting works of Pope John Paul II).

> A culture . . . constitutes the response that a given people have to these fundamental questions, a response that is constantly being revised and worked out over time. It is expressed not only through the customs and traditions of a people, but through their language, history, art, commerce and politics. Indeed, "[a]ll human activity takes place within a culture and interacts with culture." . . . As such, every culture is, in essence, a normative and didactic enterprise. It indicates what is desirable and permissible within a given society. It instructs both the observer and the participant as to how they ought to act. . . . That is, a culture is a societal answer to the question of value. Every culture renders a whole series of judgments as to what is truly important in life.

For this reason, I deliberately choose "cultural environment" rather than "information environment" or "intellectual environment."[13] "Cultural" captures the contextual, contingent, and social/relational aspects of the resources that constitute the meta-environment; the resources are resources vis-à-vis their *meaning* to and among people.[14] As Yochai Benkler suggests, "[Culture] is a frame of meaning from within which we must inevitably function and speak to each other, and whose terms, constraints, and affordances we always negotiate. There is no point outside of culture from which to do otherwise."[15] In a real sense, culture itself is an environmental concept.

One of the most important differences between the natural environment and cultural environment is the degree to which cultural resources are manufactured, both by humans and by law. Natural resources typically are given. Yet this difference is easily misunderstood or taken too far. Although the natural environment is given and not made by humans, it is continuously and unavoidably affected by humans and, in a sense, made and remade and unmade with irreversible consequences through those interactions. And although the cultural environment is made by humans, it is also inherited, subject to considerable path dependencies that can have irreversible consequences, and even contingent on human interactions with the physical environment.[16]

Viewed through the infrastructure lens, two points are clear: First, the cultural environment constitutes mixed infrastructure.[17] Human beings produce an incredible

---

[13] The other adjectives seem reductionist in the sense that they "cleanse" the discussion of normative values.

[14] Frischmann 1094 (2007a). Another reason is to connect this discussion with the "movement and ideas" associated with cultural environmentalism. See *id.* See Boyle (1997).

[15] Benkler 282 (2006a).

[16] Gordon (1993) (discussing path dependencies in the cultural environment).

[17] It is nonrival and sharable. As noted, it is comprised of nonrival resources. One might argue that some resources that are integral parts of cultural systems are partially nonrival—for example, libraries or even communications infrastructure. Such an argument does not undermine the point I am making and can be integrated into the discussion. I leave this additional layer of complication aside and focus on the intangible, intellectual, cultural, and social resources that are nonrivalrously consumed.

diversity of private, public, and social goods simply by living in and interacting with the cultural environment. It would be a mistake to assume that such productivity is inevitable or natural. In some respects, some may be; one way or another, human beings experience their lives, and "[e]xperience constitutes an important intellectual resource that simultaneously relates human beings to their inherited and evolving environment(s) and constitutes a resource that may shape the intellectual environment."[18] But the fact that the cultural environment shapes both our capabilities to be culturally or intellectually productive and our beliefs and preferences regarding the exercise of such capabilities is a reason to pay close attention to the dynamic relationships between users and the cultural environment. Whether people are active participants in intellectual, cultural, and social resource production, actively shaping the cultural environment, or passive consumers shaped by the cultural environment (or by those who are shaping it) depends on the cultural environment and how it is managed.[19] Different evolutionary paths are possible—for our environment and for us:

> [People] may become more aware, conscious of their (potential) roles as listeners, voters, and speakers, but also as consumers and producers, as political, cultural, and social beings, as members of communities. They may learn to be productive—or learn to want to be productive, if such desire is not simply latent. This very awareness that one can play different roles and that the environment is not fixed or fully determined by others is encouraging. It encourages participation and the development of facilitative social practices, and perhaps over time, the adoption of a participatory culture. . . .[20]

Second, the case for commons management is incredibly strong. Managing the cultural environment as a public commons maintains flexibility and maximizes the social option value of the infrastructure. At this scale and level of abstraction, commons management aims to limit both government and market shaping of the environment and our lives, plans, beliefs, and preferences. In other words, commons management is a strong default position for the cultural environment because users—autonomous individuals as well as social groups and communities—get to shape the environment and choose what to say and do and how to plan their lives, experiences, and interactions with each other

[18] Madison, Frischmann, & Strandburg 685 (2010). "Experience (or perception or observation) is not enclosed within IP regimes except when expressed and embodied in a particular qualifying form, such as a copyrightable work of authorship or a patentable invention." *Id.*

[19] Benkler 150–51 (2006a). I discuss this point further in my review of Benkler's book. Frischmann 1123–28 (2007a).

[20] See *id.* See also chapter 13 (considering how the Internet affects society and the cultural environment). Wendy Gordon makes this point well in the copyright context. Gordon (1993).

and the environment.[21] The cultural environment is spillover-rich because of the many different user activities that produce, distribute, use, and reuse public and social goods.

Support for commons as a default position at this macro level seems to be reflected in both the First Amendment, which restricts the exercise of government power to control the cultural environment,[22] and the related conception of a robust public domain, which limits private ownership, control, and exclusion over swaths of cultural, intellectual, and social resources.[23] Nonetheless, whether commons management truly *is* the default position is debatable.[24] Markets and government have played and will continue to play incredibly important roles in shaping the cultural environment. These social systems and institutions depend on and are essential to the cultural environment. As Benkler aptly describes (and critiques), the reality of our modern existence is that the industrial information economy and mass media system have had a tremendous influence on both the cultural environment and the American people.[25] Still, the case for commons management is incredibly strong; it should be the default position and government and market interventions and institutional structures should be understood as targeted exceptions. I return to this idea below in the context of intellectual property laws and my view that these legal systems should be understood as important, targeted exceptions.

The cultural environment as infrastructure has an intergenerational dimension. Each generation is blessed beyond measure with the intellectual and cultural resources it receives from past generations; each generation experiences and changes the cultural environment and passes it on to future generations. That we "stand on the shoulders of giants" is often noted to emphasize the cumulative nature of cultural or scientific progress.[26] But "the expression also reflects an understanding of intergenerational dependence: each generation is both dwarf and giant; the current generation stands on the shoulders of the past and also serves as the shoulders for the future."[27] Essentially, "shoulders" refers to the "fundamental blessings"[28] of resources preserved, created, and transmitted. While each generation faces supply-side problems (discussed below), it does so within the existing cultural environment, while also shaping the cultural environment for the future.

---

[21] Benkler (2006a) provides a richer account.

[22] Frischmann (2008b).

[23] On the public domain, see, for example, Litman (1990); Benkler (1999); Lessig (2001b); Samuelson (2006).

[24] This is tricky because the question can be approached as a matter of normative commitment—that is, does society demonstrate a normative commitment to this baseline via legal or other forms of public commitment?—and the question can be approached empirically—to what degree are markets and/or government actually shaping the cultural environment?

[25] Benkler (2006a).

[26] Scotchmer 29 (1991).

[27] Frischmann & McKenna (2011); Gordon (1993).

[28] Lincoln, Lyceum Address (referring to "fundamental blessings").

## B. Economic Characteristics of Intellectual Resources

The economic characteristics of intellectual resources are complex. We discussed most of the basic economic issues in chapter 3: intellectual resources are public goods, often a form of capital, and often the source of various types of externalities. An added layer of complexity not discussed in chapter 3 is the fact that intellectual resources are part of cultural, intellectual, and social progress, and thus a part of our complex and evolving cultural environment. In this section, I explain the supply-side problems that flow from the public goods nature of intellectual resources, and then I discuss the added layer of complexity.

### 1. SUPPLY-SIDE PROBLEMS

Intellectual resources face well-known supply-side problems, common to public goods, discussed in chapters 3 and 8. First, the inability to (cheaply) exclude competitors and nonpaying consumers (free riders) presents a risk to investors perceived ex ante (prior to production of the good), and this risk may lead to undersupply. This problem is a function of (non)excludability. Second, even if exclusion is feasible at low cost, nonrivalry suggests that markets still undersupply various intellectual resources.

#### a. (Non)excludability

Recall that (non)excludability is not a fixed or inherent characteristic of a resource; the costs of exclusion vary considerably with technology and context. In the absence of some institutional solution, there would be a significant underinvestment in *some types* of intellectual resources because of the risk that competitors would appropriate the value of the resources and undermine the ability of investors to recover their costs.

Whether private incentives are in fact inefficiently suppressed by this potential misappropriation risk depends on the type of investment, the intellectual resource in question, and the particular context. Many intellectual resources are not subject to this particular supply-side concern; we generate the resources without being disabled by concerns over misappropriation. For example, human experience generates substantial intellectual resources naturally. Paying attention to and recording one's experience entails fixed costs that we may choose to avoid. This highlights a distinction between creating the intellectual resource and identifying it as a resource, and converting the intellectual resource from a purely intangible creature of human intellect—referred to as tacit knowledge—to a recorded or fixed form that can more easily be preserved and shared. Underinvestment in these extra steps may constitute a supply-side problem that warrants attention.

In many situations, people make investments because the expected private benefits exceed the fixed costs, regardless of whether or not others free ride. Appropriating benefits through market exchange of the intellectual resource or some derivative product may

not be *relevant* to the investor. For example, we engage in many intellectually productive activities because participation itself provides sufficient private benefits.[29] Participation can be fun, intellectually stimulating, educational, or service-oriented, among other things. Participation may not be effortless or free; it may require substantial investment. Regardless, the private value derived from participation may be sufficient, and external benefits conferred to others that use or consume the output (i.e., the intellectual resource) may be irrelevant to incentives to invest. Even if those benefits could be internalized, such internalization could potentially decrease incentives to invest and prove quite costly.[30]

In many situations, people create, invent, and innovate because the anticipated returns from their own use of the results are sufficient to justify the investment. There is a rich literature on user innovation that demonstrates quite clearly how many significant innovations result from users seeking to solve their own particular problems, needs, or curiosities.[31] The key point is that similar to folks who participate in intellectually productive activities because of the direct benefits of participation, people often engage in such activities because the results (outputs) may be beneficial for themselves, and they do so without disabling concern over free riding.

In some contexts, people produce intellectual resources and welcome free riding by others. Sharing intellectual resources can be a viable strategy for increasing returns generated through other means. Benkler describes a bunch of different examples, ranging from lawyers who write articles to attract clients to software developers who share software and make money by providing services to users.[32] Sharing may help attract attention, build a reputation, lead to reciprocal sharing, and so on.

Finally, even where free riding is a concern and appropriating benefits through market exchange of the intellectual resource or some derivative product is relevant to investment decisions, self-help mechanisms, such as lead-time advantages and barriers to entry, may provide sufficient protection against free riding by competitors to support the investment. In some instances and in certain industries, self-help mechanisms are preferred for gaining a competitive advantage. Surveys of R&D managers show that factors such as securing lead-time advantages, increasing learning, developing complementary products, and ensuring secrecy are more relevant to incentives to invest in R&D than any perceived

[29] Madison, Frischmann, & Strandburg (2010); Frey 35 (2008).

[30] As discussed in chapter 3, the transaction and institutional costs may be significant, and internalization may affect other dependent markets and activities. The shift to internalization may even affect the attractiveness of the activity itself; participation in the activity may be less attractive when commercialization of the outputs occurs. Pink 37 (2010); Amabile 17 (1996); Benkler 298 (2006a).

[31] Von Hippel (2005).

[32] Benkler 42–46 (2006a).

ability to secure traditional intellectual property protection.[33] The extent of the benefits that inure from such self-help mechanisms vary by industry and depend on a multitude of external factors, including technology-enhanced access and copying opportunities. Such mechanisms are imperfect and do not suffice for many types of intellectual resources, but when relevant, they can be quite important and should be evaluated in comparison with each other and alternative institutions.

Intellectual property laws are a prominent but by no means exclusive means of addressing the supply-side problem where free riding is a concern and appropriating benefits through market exchange of the intellectual resource or some derivative product is relevant to investment decisions. Consider patent law.[34] In the absence of patent law, there would be a significant underinvestment in some types of inventions because of the risk that competitors would appropriate the value of the inventions. Granting inventors patents lessens the costs of exclusion, raises the costs of free riding, encourages licensing, and, as a result, makes a greater portion of the surplus generated by the invention appropriable by the inventor. The exclusivity provided by patent law does more than affect investment in invention, however. Patents affect the supply-side functioning of markets for inventions as well as markets for derivative products, including additional improvements, innovations, and commercial end-products. The reward, prospect, and commercialization theories of patent law take patent-enabled exclusivity as the relevant means for fixing a supply-side problem—the undersupply of private investment in the production of patentable subject matter or in the development and commercialization of patentable subject matter that would occur in the absence of patent-enabled exclusivity.[35] The theories differ largely in terms of where in the supply chain patent-enabled exclusivity is needed, and in terms of the degree of control/exclusivity needed. In reality, these needs vary by industry and context, giving each of the theories some support. As the next section explains, intellectual property laws also set boundaries around intellectual resources in a manner that reduces transaction costs and reduces information costs. This boundary-setting function creates legal "things" that can be more easily subject to market exchange.[36]

The supply-side benefits provided by patent law are not costless. The appropriation of a greater portion of the surplus presumes an increase in price. Absent exclusivity,

---

[33] Mansfield (1981); Mansfield 174 (1986); Levin et al. 795–97 (1987); Cohen, Nelson, & Walsh table 1 (2000); Barnett 1257–69 (2004).

[34] A similar story can be told with respect to copyright law, although in many contexts the emphasis shifts to supply-side problems further down the supply chain than authorship, i.e., facilitating the development, commercialization, and distribution of works of authorship.

[35] Landes & Posner (2003) (reward); Kitch (1977) (prospect); Kieff (2001) (commercialization); see also Ghosh 1353–57 (2004) (connecting prospect and commercialization theories with the theoretical work of Demsetz).

[36] Madison (2005).

competitive distribution and use of the invention would drive price to marginal cost (zero), in which case consumers would capture the full consumer surplus. When relevant, patents may enable pricing above marginal cost and, as result, introduce deadweight losses. Keep in mind, however, that the magnitude of the deadweight losses depends on both the strength of the legal rights conferred and the market conditions. (A patent might enable average cost recovery because the patent owner can exclude other competitors from free riding on sunk fixed costs and thus push competitors to sink their own fixed costs in developing a competitive substitute, but the patent need not, and typically does not, confer market power.) There are transaction and administrative costs to consider as well. But this simple explanation of patent law reveals the basic trade-off between static and dynamic efficiencies; we tolerate some deadweight losses along with transaction and administrative costs to mitigate the supply-side risk of underinvestment.

### b. Nonrivalry

Addressing the excludability problem for intellectual resources through intellectual property or other means does not eliminate the nonrival nature of the resources or ensure efficient market provisioning.[37] Some scholars have suggested that private property rights convert the public good into a private good, but this is not correct.[38] As chapter 3 discussed, (non)excludability should not be confused or conflated with nonrivalry. It may be the case that exclusion can prevent sharing, but that in no way affects the capacity of the resource or the corresponding option to share among many users. Consider three important implications.

First, nonrivalry enables sharing and an extra degree of freedom in managing or allocating the intellectual resource. For purely consumptive ideas,[39] this prompts the classic trade-off between static and dynamic efficiencies—for an existing idea, open sharing

---

[37] Lunney 994 (2002) (quoting Samuelson 387 (1954)).

[38] Samuelson 335 (1958):

> You might think that the case where a program comes over the air and is available for any set owner to tune in on is a perfect example of my public good. And in a way it is. But you would be wrong to think that the essence of the phenomenon is inherent in the fact that the broadcaster is not able to refuse the service to whatever individuals he pleases. For in this case, by use of unscramblers, it is technically possible to limit the consumptions of a particular broadcast to any specified group of individuals. You might, therefore, be tempted to say: A descrambler enables us to convert a public good into a private good; and by permitting its use, we can sidestep the vexing problems of collective expenditure, instead relying on the free pricing mechanism. . . . Such an argument would be wrong. Being able to limit a public good's consumption does not make it a true-blue private good. For what, after all, are the true marginal costs of having one extra family tune in on the program? They are literally zero.

[39] Chapter 3.

generally maximizes social welfare[40] because the marginal cost of sharing with someone is zero, but such sharing may have consequences for dynamic efficiency if it lessens investment incentives. Exclusion does not eliminate this trade-off; it simply provides the entity with the capability to exclude with the opportunity to decide whether or not to do so. For ideas, nonrivalry prompts a more complicated trade-off among static efficiency and various types of dynamic efficiencies.[41] In sum, the private and public opportunity to leverage nonrivalry remains an important economic consideration, even when the costs of exclusion are minimal.

Second, demand-measurement problems still lead to undersupply by markets even when the costs of exclusion are minimal.[42] There are two notable demand-measurement problems, one focused on "optimality conditions" and difficulties in accurately measuring consumer preferences, and one focused on externalities. I discussed both extensively in previous chapters. With respect to the first, Paul Samuelson noted that a second type of free riding occurs when consumers strategically misrepresent their true preferences in the hope that other consumers will bear a greater proportion of the costs. This problem, however, is independent of exclusion. The same is true of demand-measurement problems associated with externalities. The bottom line is while exclusion facilitates market provisioning, markets still systematically undersupply *some* public goods because market demand fails to accurately reflect social demand. I revisit the demand-side issues in the next section.

Third, reducing exclusion costs fixes an important supply-side problem and brings the supply-side analysis of market provisioning of intellectual resources in line with the discussion in chapter 8. Specifically, while natural monopoly is less often a concern, the cost structure of supply can impact incentives to invest and impose deadweight losses during fixed cost recovery. Excludability does not eliminate this issue either. As chapter 8 explained, the relevant economic baseline for evaluating the sufficiency of market incentives to invest should be average cost recovery. Sufficient incentives to invest depend on

---

[40] Caveats: First, open sharing does not mean force-feeding. People who want the idea can get it, but no one is forced to consume it. Second, I am assuming that the idea is beneficial rather than harmful. Third, I am assuming away negative network effects, for example, where my consumption of the idea makes it less valuable to you.

[41] As chapter 8 discussed, the conventional characterization of access vs. incentives as static efficiency vs. dynamic efficiency is often a gross and distorting oversimplification.

[42] If exclusion is coupled with *perfect* price discrimination, the first demand manifestation problem goes away. On why I reject that red herring, see chapter 6. Lunney explains:

> The literature also establishes that we can achieve a Pareto efficient outcome in the production of the public good by enabling perfect price discrimination with respect to the public good. In this context, perfect price discrimination creates personalized markets for the public good, where each consumer's consumption of the public good becomes a distinct commodity with its own market and its own price. If it could be achieved, the resulting equilibrium, known as a Lindahl equilibrium, would essentially convert the public good into a private good and ensure a Pareto efficient outcome.Lunney 451–52 (2008).

an expectation of recovering total costs, including a competitive return on capital investment. The cost structure suggests that incentives to invest will be insufficient and undersupply will result, unless pricing above marginal cost and (at least) approximating average cost is sustainable. Without exclusion enabled by intellectual property or other means, it might be impossible for suppliers to recover their average costs because free-riding competitors would drive prices to marginal cost. Exclusion can enable sustainable average cost pricing and competition, in a sense facilitating markets. Enabling average cost pricing does not ensure actual cost recovery, however. As chapter 8 discussed, there are a number of practical obstacles to effectively implementing average cost pricing. Moreover, competition and innovation can jeopardize cost recovery, for example, when a new entrant figures out a way to compete with lower fixed costs. If exclusion is limited in scope to actual misappropriation (in essence, free riding on the fixed cost investment of the first entrant), then whether the first entrant is capable of recovering its costs will depend on the fixed cost investments that others must make to enter the market as well as lead-time advantages and other possible barriers to entry.[43] On the other hand, exclusion also can enable monopoly pricing and eliminate competition.

What exclusion enables depends on the strength and scope of exclusion and the market context. The legal right to exclude can be narrowly or broadly constructed along various dimensions. For example, it can be limited to actual copying of an entire intellectual work, broadened to block copying of parts of the work, broadened to block similar but not identical copying, or broadened beyond instances of copying to block independent creation, among other things. It can also vary in other dimensions such as duration, the strength of remedies, and so on. At the extreme, government could grant monopoly franchises with legal entry barriers. The point is that exclusion can vary in strength, scope, and market impact.

[43] People sometimes emphasize the magnitude of fixed costs. In many cases, this doesn't really matter so long as a second comer would have to sink the same amount. High fixed costs may actually be a decent barrier to entry, provided that misappropriation is precluded and average cost pricing is feasible. What seems to matter in such circumstances is the rate of fixed-cost-reducing innovation—whether a second comer can figure out a way to enter more cheaply. Of course, this is true in all sorts of markets. Certainly in some cases, incredibly high fixed costs may exceed capital constraints for any single firm, but that raises a different problem altogether.

A difficult supply-side question[44] to confront is whether patent or copyright law should do more than address *actual* free-riding risks.[45] If patent and copyright laws aim to facilitate competitive markets for intellectual resources and derivative products, a narrow focus on such risks would be appropriate. However, if the laws aim to induce investment in intellectual resources above and beyond what competitive markets would provide, a broader focus on conveying market power and the ability to appropriate supracompetitive returns might be appropriate. If the latter objective is chosen, however, then one

---

[44] I raise this question because it gets insufficient attention in intellectual property scholarship, and, as a result, some unfortunate assumptions/overstatements are made. I highlight two:

- First, a common overstatement in intellectual property discourse suggests that intellectual property laws *create* incentives to invest. Intellectual property laws do not create incentives exactly. Generally, incentives to invest exist independently of intellectual property laws. The motivations briefly described in the previous subsection constitute incentives to invest, and those motivations—whether driven by the value of participation, prospective use of the output, or prospective appropriation of value through market exchange—are not constructed or "created" by intellectual property laws. Rather, intellectual property laws address risks that may distort markets and deter some people from doing (investing in) what they would otherwise be inclined to do (invest in)—in a sense, intellectual property laws assist in the construction of a market. To the extent that the risks are irrelevant and do not distort markets, then the justification for intellectual property is greatly diminished from an economic perspective. (Some might argue that intellectual property rights are still necessary to reduce transaction costs and facilitate coordination, but this argument is significantly diminished where the risk of free riding is irrelevant because it is not clear what intellectual property offers above and beyond what traditional means for coordination already provide.) By addressing free-riding risk when it is relevant, intellectual property (re)aligns incentives to invest, but it seems odd to say that in doing so the law *creates* those incentives. One reasonably could say that intellectual property laws counteract the particular disincentive associated with misappropriation risks and, in that very limited sense, create incentives.
- Second, the conventional economic explanation of intellectual property sometimes slips into a story about temporary monopoly or market power that allows intellectual property owners to extract monopoly rents. This story would seem to support the argument that intellectual property rights create incentives because the prospect of extra-market returns would induce investment above and beyond what would otherwise exist in a competitive market. But recall our discussion in chapter 8 about the appropriate economic baseline for sufficient incentives to invest; the baseline is average cost recovery. To the extent that incentives to invest hinge on prospective appropriation of value through market exchange, the market creates those incentives rather than intellectual property law; providing exclusion by legal means lessens the otherwise disabling costs of exclusion associated with public goods provision and may reduce information and transaction costs associated with appropriation of value through market exchange, but absent additional justification, there is little reason to put a thumb on the scale and provide further inducement to invest via the prospect of extramarket returns. I am not suggesting that additional justifications do not exist; I discuss some in the following sections. But those justifications are complicated and cannot be based exclusively on the risk of free riding or mere evocation of public goods. Moreover, to the extent that society wishes to provide such additional inducement, a comparative institutional analysis would be required.

[45] I emphasize *actual* free riding *risks* to remind you that free riding does not always present a risk to investment because alternative means of exclusion may exist and alternative motivations may provide sufficient incentives irrespective of free riding. See chapters 3, 8; Lemley (2005); Frischmann & Lemley (2007); Liivak (2010); Le 32 (2004).

would have to both justify the need for extra inducement (Why put a thumb on the scale in favor of investments in intellectual resources rather than other types of investments? Is the increase in deadweight losses worth the gain?) and explain from a comparative institutional standpoint why intellectual property laws are the preferred institution for making this social investment—why intellectual property rights rather than government subsidies, a prize system, or other alternatives.[46] Patent and copyright law differ substantially in their institutional design—for example, patent provides a stronger right to exclude than copyright but for a much shorter duration—and it might be argued that patent law is more directly attuned to the latter objective. Note, however, that most intellectual property rights do not in fact convey market power that would allow a supplier to sustain prices above competitive levels, which we might expect to gravitate toward average cost pricing over the medium to long run. In some cases, market power does arise, but it is debatable whether such market power is attributable to the granting of the intellectual property right, the success of the innovation, or other context-specific factors.[47]

An important implication of the cost structure of supply is that market provisioning involves deadweight losses and the magnitude of those losses may be quite significant given the high ratio of fixed cost to marginal cost. There are a host of deadweight mitigation strategies, ranging from price discrimination to government prizes. Even the limited duration of intellectual property rights can be understood as deadweight mitigation strategy. I return to this issue in the context of intellectual infrastructure.

## 2. Intellectual Resources and Activities, Products and Processes

The previous section focused on the basic supply-side problem; this section focuses on the added complexity associated with "the other crucial quirkiness . . . that information is both input and output of its own production process." [48] As noted, this effect is interesting and complex because it reveals necessary dependence among generations, but there is more to it than that. It implicates the cumulative, dynamic, and evolutionary nature of progress in intellectual-cultural systems, or, more broadly, in the cultural environment. I examine a few distinct but related points that often are conflated or ignored in discussing this quirkiness.

Benkler focuses on how the "on the shoulders of giants" effect makes "property-like exclusive rights less appealing" because it increases the deadweight losses from pricing

[46] Fisher 200–04 (2004); Madison, Frischmann, & Strandburg 685 (2010). Note that alternatives might include support for infrastructures that enable nonmarket production.

[47] FTC ch. 3 (2003) (collecting evidence that "issues of fixed cost recovery, alternative appropriability mechanisms, and relationships between initial and follow-on innovation" vary by industry); Burk & Lemley 1577–1589 (2003) ("Recent evidence has demonstrated that this complex relationship [between patents and innovation] is . . . industry-specific at each stage of the patent process").

[48] Benkler 37 (2006b).

above marginal cost of zero by making productive use of the nonrival resources more costly. He notes: "Today's users of information are not only today's readers and consumers. They are also today's producers and tomorrow's innovators."[49] Simply put, users are both consumers and producers. Obviously, I agree (given the discussion of this general problem in earlier chapters). The fact that many intellectual resources are a form of nonrival capital that supports production of even more nonrival capital suggests the possibility of increasing returns to investing in such resources and leveraging nonrivalry.[50]

Yet we take the "on the shoulders of giants" effect for granted.[51] For example, we often take for granted the intellectual or cultural backdrop within which and on which we (and others) build; this may be due to a romantic notion of authorship, an inflated sense of self, or any number of things.[52] Similarly, we often take for granted the various intellectual outputs that emerge from our experience and engagement with the cultural environment; we only have so much time and attention. Regardless, we use, make, and reuse intellectual resources continuously in our lives. This seemingly trivial observation has some interesting implications. First, we need intellectual inputs to be intellectually productive and to make intellectual progress in our lives. Second, the intellectual resources to which we have access will shape the intellectual outputs we are capable of producing as well as our beliefs and desires about what to produce; in a sense, they shape who we become (our beliefs, knowledge, preferences) as we engage with the environment. Third, each producer and producer's output is thus dependent or contingent on various inputs. "In order to write today's academic or news article, I need access to yesterday's articles and reports. In order to write today's novel, movie, or song, I need to use and rework existing cultural forms, such as story lines and twists."[53] In a sense, this is a more micro-way of

---

[49] *Id.*

[50] This is a key dimension to Romer's growth theory. Romer (1996); Ochoa 10–15 (1996). Although I refrain from discussing macroeconomics much in this book (to keep some limit on the scope!), there is an interesting connection between these features of intellectual resources and processes/activities and some of the new growth models. Schmidt 11 (2003); Romer (1996).

[51] To the extent that this effect is taken seriously in economic and legal scholarship, attention is devoted to the relationship between two stages, the first- and second-generation producers, the pioneer and improver. For example, Suzanne Scotchmer has focused on this effect. She emphasizes the importance of licensing intellectual property between first- and second-generation inventors, and adequately compensating and maintaining investment incentives to both stages of inventorship, given that many products are the result of numerous improvements on previous inventions. Scotchmer (1991); Green & Scotchmer (1995); Scotchmer (1996); Lemley (1997); Merges & Nelson (1994); Merges & Nelson (1990). See also Technological Infrastructure Policy: An International Perspective at 8 n. 2 ("Cumulative forms of knowledge are those in which today's advances lay the basis for tomorrow's, which in turn lay the basis for the next round. The integrative aspect of the production of knowledge means that new knowledge is selectively applied and integrated into existing systems to create new systems.").

[52] Many scholars have discussed this point. See, e.g., Litman (1990); Boyle 122 (2010); Lessig 9–13 (2001b); Vaidhyanathan 117–48 (2001). See also Lemley (forthcoming 2012).

[53] Benkler 37 (2006b).

making the point I made earlier with respect to the cultural environment. As Julie Cohen suggests, we are situated within the cultural environment, shaping it while being shaped by it.[54]

Continuous situated engagement implies a stream of input-output relationships (i.e., input → output/input → output/input . . .). In many contexts, it may not be worth the effort to pay attention to the continuous streams of relationships. Surely, we do not need to acknowledge and consider each incremental addition associated with sensory experience or thinking. Instead, we may conflate many input-output relationships into a *process* (activity or practice) and pay attention only to particular outputs that are worthy of attention.[55] Note that such a conflation begs for deeper interrogation. Why would we do this? How do we choose to distinguish between resources and processes? When do we decide to pay attention to the outputs? When are they, and how do they become, meaningful or worthy of attention?[56] When is something an input, an output, both, or an aspect of a process? And so on. This is not the place to address such questions, however. I make two observations and then push on:

- It is common to talk about intellectual resources as identifiable, discrete things with known properties and boundaries. The very notion of a "resource" or "public good" implies such features. But this is a significant oversimplification. Intellectual resources often have a dual nature—creation, invention, and innovation may be resources and activities. Consider basic research: Is it a thing—a result, an input, an output, both—or is it a process or activity that one engages in? It is both, right? Maybe this seems like a semantic point, but isolating one from the other (product from process) for purposes of the law, economic analysis, or just discussion loses something quite valuable.
- Can we "discretize" cumulative intellectual processes of creation, invention, and innovation in a manner that makes analytic sense? We try to do so regularly within copyright and patent law, but are we truly granting patents and copyrights over discrete outputs—over discrete "things"? When we are dealing with streams of input-output relationships that may or may not culminate in a consumer good, it can be difficult to isolate the "thing" we might identify as *the* invention or work of authorship, much less the intellectual contribution made by the person claiming patent or copyright.[57] We recognize and enforce (artificial)

[54] Cohen (2005) focuses on the situated user and the importance of the dynamic relationships between creators and context. Her situated user "engages cultural goods and artifacts found within the context of her culture through a variety of activities ranging from consumption to creative play." *Id.* at 370.

[55] Frischmann (2000).

[56] Benkler explores this in the context of communications. Benkler (2006a).

[57] I discuss this issue further below. On the difficulties and importance of delineating "things" in a variety of contexts, see Madison (2005).

> boundaries for purposes of constructing property rights and facilitating exclusion, coordination, and market provisioning, but our focus on "things" (inputs, outputs, resources, goods, and so on) often obscures the continuity and complexity of the system.

Even if we reduce the number of input-output relationships we are willing to entertain, we must acknowledge that intellectual progress involves a stream of such relationships, and this requires acknowledgment of a potential stream of spillovers, or what we might refer to as *cascading spillovers*. This is the case even if we assume a simple string of single public good input-output relationships, where a single public good is produced at each stage. In reality, each stage of production may involve multiple input and outputs, each of which can be used productively to produce different outputs and potentially support different production paths by many people. Many intellectual and cultural activities yield social goods as well, in which cases the diffusion of a different set of externalities cascade as well.

The conventional model of intellectual production represents progress in a linear fashion, for example, from basic research to applied research and finally to commercial application; or, alternatively, from idea conception to invention to commercial development; or something similar. Linear models are intuitive and qualitatively appealing because many economic and social policy questions that follow seem to have straightforward answers.[58] For example, the government should support basic research as a form of public goods production; the basic research pool should supply inputs for applied research; private firms should step in at some point and bring the benefits of research results to the public through commercialization.[59] Yet the linear model is not an actual scientific model of innovation or intellectual progress. Rather, as Benoît Godin explains, a host of different actors—scientists seeking funding, economists advising government agencies—constructed the linear model of innovation to classify research activities, establish a connection between basic and applied research and eventually commercial activities, and advance political and other agendas.[60] Godin explains that the simple three-stage "basic research → applied research → development" model became standardized when official government statisticians appropriated the three-stage model as a means for classification of research to aid in statistical categorization, measurement, and quantitative analysis.[61] Yet the linear model has been roundly criticized and rejected.[62] As Nathan Rosenberg

---

[58] Although many trace the linear model to Bush (1954), Godin suggests that "[o]ne would be hard-pressed . . . to find anything but a rudiment of this model in Bush's manifesto." Godin 639 (2006).

[59] Frischmann (2000).

[60] Godin (2006).

[61] *Id.* See also OECD 12 (1962).

[62] Kline & Rosenberg (1986).

claimed in 1994, "Everyone knows that the linear model of innovation is dead."[63] Yet as Godin shows, the linear model remains intact in the discourse despite its many criticisms. He observes that alternative models have struggled to replace the linear model because they pose more difficult measurement issues, and "with their multiple feedback loops look more like modern artwork or 'a plate of spaghetti and meatballs' than a useful analytical framework."[64]

Intellectual production processes, and intellectual progress more generally, are often nonlinear, multidirectional, stochastic, full of feedback loops, and difficult to model.[65] There are various nonlinear innovation models that incorporate dynamic interactions between different types of research and even nonresearch activities as well as the background cultural environment within which such interactions take place.[66] For example, the "Chain-Linked Model," developed by S. J. Kline, incorporates feedback loops between research and the "existing corpus of knowledge" and emphasizes the importance of various different activities, procedures, and external influences that play a role in innovative progress; there are multiple paths, feedback loops, and various actors.[67] As Kline and Rosenberg point out, the linear model's omission of feedback loops, learning from "shortcoming and failures," and other features renders it incapable of dealing with radical and incremental innovation.[68]

The dynamic nature of progress often leads to unexpected spillover effects. For example, an idea developed in one sector may lead to beneficial progress in another unrelated (or marginally related) sector.

> The practice of science is becoming increasingly interdisciplinary, and scientific progress in one discipline is often propelled by advances in other, often apparently unrelated, fields. For example, who would have thought that nuclear physics research (the study of the inner workings and properties of the atomic nucleus) and data gathering techniques developed for experiments on elementary particles (quarks and such) would lead to a device that has advanced the boundaries of biomedical research and health care? Yet both of these lines of inquiry led ultimately to Magnetic Resonance Imaging (MRI), a tool now used in laboratories and hospitals around

[63] Rosenberg 139 (1994).

[64] Godin 639 (2006). Frankly, this seems like yet another example of looking for what can be measured more easily rather than what actually matters, or, to return to the old joke, looking for one's lost keys only under the lamppost.

[65] Dreyfuss (2010); Godin (2006); Knudsen 13, 24 (2003) (examining the policy implications of self-reinforcing processes and traps such as feedback loops in intellectual progress).

[66] Padmore & Gibson (1998); OECD (2000).

[67] Kline & Rosenberg (1986); Kline (1991a); Kline (1991b).

[68] Kline & Rosenberg (1986). To be clear, many scholars that study innovation systems recognize these complexities.

> the world both to conduct basic biological research and also to diagnose illness. Such cross-over between fields is yet another example of the unexpected payoffs that can come from basic research.[69]

There are countless examples in science, technology, and innovation, but these phenomena are equally relevant in cultural systems too. James Boyle presents one such example of cross-genre evolution in the music field: the story of an amateur hip-hop song entitled "George Bush Doesn't Care about Black People," released in 2005, that sampled Kanye West's song "Gold Digger" from the same year and was named after Kanye's outburst criticizing former president George W. Bush for his response to Hurricane Katrina. West's song had, in turn, sampled the 1950s R&B/soul song "I Got a Woman" by Ray Charles, which, it turns out, was a rewording of the Christian hymn "Jesus Is All the World to Me," penned in 1904.[70] In a more direct way, music sampling reflects the cross-pollination of musical genres, albeit through conscious appropriation rather than organic development.[71] Literature is rife with examples of similar developments. Genres blend into each other and form new hybrids whose existence is sometimes fleeting, sometimes stable: Twentieth-century American musicals spawned cowboy musicals and gangster musicals like *Guys and Dolls*;[72] the cult classic *Blade Runner* drew thematic elements from pulp genres such as film noir and science fiction to create a subgenre of its own.[73] Cultural systems affect each other in complex and unpredictable ways; for example, geographically driven changes in the interactions of ethnic groups affect literary traditions, art, and architecture.[74]

The cultural environment and its constituent innovation, science, culture, knowledge, and other systems are dynamic evolutionary systems. Since how and what direction the systems, environment, and consequently society evolve are not predetermined or inevitable, institutions and social policies matter considerably; the cultural environment we construct and sustain reflects deep normative values.[75]

---

[69] Staff of House Comm. on Science (1998); see, e.g., Nelson 459 (1982) (discussing the spillover effects of government research motivated by national security interests into other applied areas).

[70] Boyle 122–25 (2010).

[71] McLeod & DiCola (2010) (music sampling).

[72] Schatz (1981) (describing the hybridization of film genres); Schatz 44 (1977); Altman (1989).

[73] Doll & Faller (1986) (describing the literary and motion picture genre cross-pollination that led to the making of *Blade Runner*).

[74] Miller xvii (2007).

[75] I am tempted to digress into a discussion of how we have lost sight of values in our focus on perfecting the means to achieve optimality. What do I mean by this? Utilitarianism and utilitarian economics has become the dominant mode of analysis, for various reasons, and a host of normative values are either forced into commensurability boxes to be traded off against each other within the utilitarian economic framework or are simply ignored. Utilitarian economics may be the analytical framework best able to provide answers to policy questions, but that does not mean the answers are correct or the best indicator of what society actually wants.

The dynamic, nonlinear, and multidirectional nature of these processes/activities involves considerable uncertainty, and this can be daunting and possibly viewed as something we hope to control, diminish, or eliminate over time. Yet when coupled with the nonrival nature of intellectual resources, it suggests considerable *social opportunity*—the opportunity to leverage nonrivalry. The nonrival nature of the resources means that intellectual capital generated at different points in the "stream" may flow along many paths, potentially being used simultaneously by different people in different settings as an input into multiple intellectual-cultural processes. There are a variety of obstacles to the free flow and use of intellectual capital (e.g., limited absorptive capacity, education, or capabilities to productively use the resources),[76] and in particular contexts, there are good reasons to restrict the free flow and use of intellectual capital (e.g., to prevent misappropriation and protect supply-side incentives to invest). My point is that in light of nonrivalry and the "on the shoulders of giants" effect, the social opportunity deserves recognition and further attention.[77]

I conclude this section by revisiting the point I made in the appendix to chapter 3. The complex, dynamic, nonlinear, and multidirectional nature of intellectual progress and the prevalence and variety of external effects in the cultural environment suggest that focusing on optimality conditions may be a red herring (or worse). We are inevitably in what economists call a second-best world because of the incredible number of incomplete and missing markets in the cultural environment.[78] Rather than focus on achieving optimal government or market selection of public good investments, society is likely (much) better off focusing on indirect interventions that (a) support public capabilities to participate in intellectual-cultural activities and (b) aim to lower the costs of public goods production for a wide range of public goods while (c) maintaining flexibility in the opportunities available to potential participants. In my view, this would better leverage nonrivalry, facilitate progress along many paths, and sustain a spillover-rich cultural environment in which and with which members of society are capable of interacting productively. This shift in focus has important implications for the subsequent discussion of intellectual infrastructure (section C) and for an appreciation of intellectual property laws as semi-commons arrangements (section D).

---

But failing to question or deal deliberately with the underlying normative values effectively deals with them. In any event, this is simply too big a digression. See, e.g., Frischmann & McKenna (2011). For an excellent discussion of a range of normative values within the liberal tradition, see Benkler (2006a).

[76] Bontis 134–35 (2005) (discussing this issue and listing strategies such as improving tertiary education, implementing R&D policies, encouraging tourism, and hosting international conferences, as means to promote the accumulation of intellectual capital assets); Albert & Bradley 79 (1997) (describing barriers to the flow of intellectual capital).

[77] Benkler (2006a) explores this social opportunity in the context of commons-based peer production.

[78] Lipsey & Lancaster (1956).

Here I briefly note what this would mean for intellectual property laws: First, the laws should focus on misappropriation risks with an aim to facilitate average cost recovery and competition, rather than market power or monopoly; to the extent that certain areas warrant subsidies, then targeted subsidies seem more direct and less distorting than adjusting the legal system for all areas, and such subsidies can be directed at infrastructural investments in the targeted area—for example, basic research in biotech or even directly funding clinical trial system. This first point suggests that exclusion is important but should be limited in scope. Second, in addition to facilitating exclusion, intellectual property systems should aim to reduce information and transaction costs because such cost reductions would apply to a wide range of public goods investments. This can be done in a manner similar to traditional property law systems, by providing recordation, registration, dispute resolution, and so on. Also, given the multitude of intellectual property owners, private ordering solutions to collective management problems should be facilitated as well. Third, and related to the first point, it should be understood that the division of surplus often has efficiency consequences (rather than mere distributional or equity consequences) because consumers are often productive users, even if their productive use does not immediately generate a marketable good. This has consequences for a variety of economic issues in intellectual property law. Fundamentally, and in stark contrast with the conventional economic perspective, intellectual property systems should be understood as exceptional, targeted interventions that construct semi-commons and sustain a spillover-rich cultural environment.

## C. Intellectual Infrastructure

### 1. Applying the Criteria to Delineate Intellectual Infrastructure

Applying the infrastructure criteria to intellectual resources delineates a broad set of resources that create benefits for society primarily through the facilitation of downstream productive activities, many of which generate spillovers. The definition of infrastructure can be reduced as follows to fit intellectual infrastructure: *nonrival input into a wide variety of outputs*. This seems incredibly capacious. Like environmental infrastructure, intellectual infrastructure can be identified and analyzed at various levels of abstraction, ranging from the meta-environment itself to a discrete general-purpose input, such as a basic idea, to a specific expression that has broad communicative power and social meaning. The resource set is much more highly and diversely populated than traditional infrastructure. Each of the following categories, for example, contains innumerable examples:

- Basic research
- Infrastructural ideas

- General-purpose technologies[79]
- Languages

Rebecca Tushnet once told me that the infrastructure concept seemed to have a fractal nature when applied to intellectual resources because you could identify infrastructure at various scales and observe repeating patterns, similar characteristics, and so on. She suggested the concept seemed to apply too easily and to too many different resources, and as a result, it seemed to lose its usefulness. While I agree with her observations, I reach a different conclusion. It seems to me that the dynamic, cumulative and interactive features of cultural-intellectual resources and practices may mean that more cultural-intellectual resources potentially function as infrastructure. Keep in mind that the infrastructure criteria describe functional relationships and are only part of the broader resource management question. That the functional relationships repeat at different scales means that choosing the appropriate scale for evaluating resource management options and trade-offs is critical. Different types of rules may be appropriate at different scales; for example, a default open-access-style rule (ex ante, broadly applicable) may be appropriate for intellectual infrastructure at high levels of abstraction, while a less demanding rule, such as an essential-facilities-style rule (*ex post*, case-specific), may be appropriate for intellectual infrastructure at low levels of abstraction.

Still, many intellectual resources do not fall within the scope of the general definition of infrastructure. A few examples illustrate this point as well as a number of complications that arise in applying the criteria.

First, consider a common construction wire nail ("common nail").[80] While the tangible nail itself satisfies the latter two prongs of the definition (input into wide variety of outputs), it fails to satisfy the first prong because nails are rivalrously consumed and cannot be managed in a way that renders consumption nonrivalrous. Consumption of the tangible good depletes the consumption opportunities of others and means that additional supply is needed to meet the demands of others. A common nail must be made for

[79] General-purpose technologies are those drastic innovations that have "the potential for pervasive use in a wide range of sectors in ways that drastically change their modes of operation." Helpman 3 (1998). These technologies often, but not always, fit into our definition of infrastructure but involve some added complications. See the appendix to chapter 2. Basically, new general-purpose technologies appear to require the development of new intermediate goods, and perhaps new infrastructure or meta-infrastructure, before the technologies can be effectively implemented. Aghion & Howitt 121 (1998). Developing these intermediate goods requires time and a critical mass of demand, which may result in a possibly painful transition as resources are taken out of production and put into R&D activities aimed at developing these new goods. *Id.*

[80] I thank Scott Kieff for this example. A "wire nail" is made from wire of the same size as the shank of the nail, by a machine that cuts the wire in even length, heads it, and points it. They are said to be stronger for driving vand less prone to splitting the wood, and are thus earned their stripes with carpenters a century ago. Kidder 324 (1899). Common wire nails are today ubiquitous, and although newer specialized nails compete for use, they "are still the most frequently used fasteners." Faherty 5.3 (1999).

each user.[81] But what about the *idea* of a common nail? Ideas are nonrival goods. The idea itself is not infrastructural, however, because it fails to satisfy the third criterion; it is special purpose rather than general purpose. The idea of a common nail is a nonrival input into the production of a *single* output—a tangible common nail, which is an input into a wide range of outputs. There is little reason to believe that demand-manifestation problems in the output markets will distort the input market.[82] Competitive markets for common nails work well in manifesting demand for production of such nails and in allocating them.

This example highlights three boundary issues that arise when applying the criteria to distinguish infrastructure and non-infrastructure. I note the issues here and then discuss each of them further in the examples that follow.

- First, it may be difficult to draw lines where there is a stream of cumulative inputs (idea of a nail → nail → range of outputs). Although this does not seem terribly significant in this particular example, it can be incredibly difficult where the stream is of the type described in the previous section, that is, streams of public goods.
- Second, it may be difficult to choose the appropriate level of abstraction. At a more abstract level, the idea of a nail—perhaps expressed as the idea of a pin-shaped fastener or the idea of a fastener that holds materials together by shear strength laterally and friction axially—may be infrastructure because the variety of outputs expands significantly with the abstraction. The common construction wire nail is but one of many different tangible embodiments of the more abstract idea.[83]
- Third, it may be difficult to choose the appropriate scope of uses. I implicitly narrowed the scope of relevant uses of the idea of a common nail when I declared it to be a nonrival input into the production of a tangible common nail. Specifically, I limited the scope of relevant uses to implementing the idea through the transformation of physical resources to produce a tangible embodiment. I maintained the same scope at a higher level of abstraction in the previous paragraph. At both levels of abstraction (idea of nail and idea of common construction wire nail), we

[81] Recall the discussion in chapter 3 of rivalrous consumption goods, raw materials, and intermediate goods. Three principles apply: (1) social welfare is maximized when a rivalrous good is consumed by the person who values it the most, and the market mechanism is generally the most efficient means for (2) rationing such goods and (3) allocating resources needed to produce such goods.

[82] Put another way, demand-measurement problems associated with externalities are not significant. Keep in mind that the demand-measurement problem identified by Samuelson—consumers misrepresent their actual preferences for the public good in an effort to free ride—might arise in the (public good) input market but does not arise in the (private good) output market.

[83] Kidder 324–25 (1899) (describing various types of nails); Faherty 5.3–4 (1999) (describing nails).

could expand the scope of relevant uses to include uses of the ideas to communicate or to develop new ideas, in addition to using the idea to produce tangible embodiments.[84] In applying the criteria, the scope issue should be resolved by considering which uses are the primary drivers of social demand.[85] Thus, one basic reason for maintaining the narrow scope for the idea of a common nail is that social demand for this idea is driven primarily by production of tangible nails and not the other possible but less relevant uses. (This is very similar to the reason given in chapter 3 for describing apples as rivalrously consumed and largely ignoring the potential nonrivalrous use of a tangible apple as an input into the production of a painting.) Such a narrow scope may be questionable, however, at the higher level of abstraction. I revisit this particular example and issue below in the discussion of ideas.

As a second example, imagine that scientists discover a drug to cure a particular disease. While the discovered knowledge is a nonrival input and thus satisfies the first two prongs of the infrastructure definition, the range of outputs may be relatively narrow, for example, curing the particular disease and perhaps some closely related research on improving the cure (e.g., to hasten the recovery). In some cases, the range may broaden to the point that the discovery is infrastructural. Putting that possibility aside, however, I would not classify the discovery as infrastructure based only on the social value of the resource. It is not the magnitude of social value associated with the resource that makes it infrastructural, although infrastructure may generate substantial social value; rather, it is the functional nature of the resource and the manner in which it generates social value that matters. As this example highlights, one of the difficulties in applying the third criterion is figuring out how wide a variety of outputs qualifies something as infrastructure. How basic or generic must the input be? At what level of abstraction do we assess genericness? There are no easy answers to these questions. To some degree, these questions permeate patent and copyright law, as I discuss below. Although it important to acknowledge and be aware of this difficulty, we do not need easy answers. There is some fuzziness at the boundaries, which is probably inescapable given the complex resource systems involved.

Regardless of whether the discovery is classified as infrastructure, there still may be a strong case for government intervention in one form or another on social welfare or other

[84] Including these additional types of uses does not involve abstraction. The scope issue arises at each level of abstraction.

[85] Alternatively, the scope issue can be resolved by looking at the institutional framework. For example, even if we consider all possible uses of the idea of a common nail, different legal regimes focus on different types of uses—patent focuses on implementation by physical transformation or tangible embodiment and copyright law focuses on communication; neither focuses directly on idea production, although both indirectly affect idea production. I discuss this further below.

grounds. If privately owned and controlled, a host of government interventions ranging from open access, essential facilities (essential medicine), price regulation, or even government expropriation might be justified by significant public health concerns.[86] On the other hand, there may be a strong case for encouraging price discrimination, particularly where markets can be segmented geographically or along similar broad market-by-market categories. Price discrimination may enable cost recovery for drugs that require incredibly high fixed cost investments in research, development, and clinical trials while significantly reducing deadweight losses and improving access to critical medicines in markets where resources are limited.[87] *This is a very complex problem*. My basic point is that the range and variety of downstream uses is often quite narrow for a particular drug; as a result, maintaining flexibility or the social option to pursue different paths downstream is less pressing, and to the extent that market failure arises, targeted approaches are probably more attractive than commons management. All of this being said, however, there are good reasons to look further upstream from the drug itself to examine the basic research and drug discovery system, which probably constitutes mixed infrastructure and is subject to substantial risk of demand-side failures and associated misallocation of resources because of distorted prioritization.[88]

Third, consider the 80386 microprocessor chip developed by Intel Corporation, commonly referred to as the "386 chip."[89] Many people might regard the 386 chip as infrastructural because it constituted a significant advance for desktop computing and was a substantial platform for innovation in software. The chip could handle 4 million operations per second and up to four megabytes of memory. It was backward compatible, meaning it worked with Intel's previous line of processors for IBM PCs and could run software written for those processors. Applying the criteria follows a similar path as the previous examples but raises a few wrinkles. The actual chip, like a nail, satisfies the latter two prongs of the definition (input into wide variety of outputs) but fails to satisfy the first prong. Each chip is a tangible, rivalrously consumed good. While the chip is durable and is not depleted or transformed upon use, it is still rivalrous. When used in an IBM PC, for example, a 386 chip cannot be used in another IBM PC. The concept of sharing a 386 chip among multiple users makes sense if the computer itself is being shared, perhaps

---

[86] Various versions of open-access regimes and open-source research collaborations have been proposed as pharmaceutical R&D models to increase access to essential medicines. Srinivas 263 (2010). Several countries have external price regulations whereby these governments set the price or reimbursement rate for pharmaceuticals and vaccines based on the prices charged in other countries. McElligott 426 n. 49 (2009). Similarly, compulsory licensing schemes have been proposed that invoke the essential facilities doctrine for public health reasons. Liu (2008).

[87] Fisher & Syed (forthcoming).

[88] *Id.*

[89] Chris Yoo raised this example at a colloquium at Wharton, and the lively discussion that followed suggested that it was worth including here.

among a community of users, such as a family or among library patrons, or among users on a network. Put this way, the computing power of the chip is sharable. In this sense, the 386 chip (or the computer) is comparable to the harvesting equipment discussed in chapter 3 (or an automobile or other types of capital equipment), in that it can be shared over time among a community of users because of its durability. There are, however, significant limits on the capacity of the resource that correspondingly limit such sharing arrangements and the opportunity to leverage nonrivalry. I do not mean to diminish the importance of this potential for sharing, but the relevant community cannot be the public at large (unless perhaps a fleet of capital equipment is shared).[90] This example highlights an underspecified part of the first criterion, specifically, what constitutes an "appreciable range" of demand. I leave that open for now and again acknowledge that there are some fuzzy boundaries around the infrastructure set.

If the 386 chip itself is rivalrous in the sense that a chip must be made for each computer, what about the *ideas or design* behind the 386 chip? Similar to the previous examples, the "idea of the 386 chip" does not satisfy the third criterion because it is special purpose in that it is used primarily to generate the 386 chip itself. As with the previous example, however, there is some room to argue because the ideas are relevant to research. This highlights another difficulty, choosing the relevant idea and the appropriate level of abstraction for analysis. The specific chip design is more easily dismissed as not infrastructure because it is primarily used to produce actual chips, but more abstract descriptions of various ideas that surround, enable, and in a sense constitute the "idea of the 386 chip" are less easily categorized. Consider, for example, the idea of backward compatibility. While relevant to describing the 386 chip and perhaps even a defining feature of the chip, this idea is separable and infrastructural, at least when stated at this level of generality. The idea of backwards compatibility applies to an incredibly wide range of products and services.

These examples test the boundaries of the infrastructure criteria and demonstrate some fuzziness, which is unavoidable. Nonetheless, focusing on intellectual resources that satisfy all three criteria for infrastructure helps to distinguish different types of intellectual resources based on the manner in which they create social value. All intellectual goods are nonrival; some are primarily valuable as consumption goods, while others are primarily valuable as intermediate goods, as intellectual capital. We are concerned with intellectual capital that is generic in nature—that can be used by many (people, firms, etc.) as an input into a wide variety of productive activities—and once we have identified such resources, we are further interested in the nature of the productive activities and whether users produce private, public, or social goods. This set of resources deserves careful

[90] On such community sharing, see Benkler (2004); Levine (2009) (describing the increasing popularity of Zipcars, fleets of cars parked in garages in various cities in North America and Europe, that members can use via an online reservation system; almost 5 million people in New York City live within a ten-minute walk from a Zipcar). Sharing initiatives in more traditional ways may accomplish similar goals, as with computer clusters at public libraries or car rental agencies.

attention because the benefits of open access (costs of restricted access) may be substantially higher than for intellectual resources that are not infrastructure.

The next section considers ideas as an example of intellectual infrastructure. Then, the following section considers intellectual property systems more generally and how those systems mediate access to intellectual infrastructure.

## 2. IDEAS

> 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.
> —JEFFERSON (1813)

> The general rule of law is, that the noblest of human productions—knowledge, truths ascertained, conceptions, and ideas—become, after voluntary communication to others, free as the air to common use. Upon these incorporeal productions the attribute of property is continued after such communication only in certain classes of cases where public policy has seemed to demand it.
> —INS V. AP

Should ideas once disclosed to the public be "free as the air to common use," as Justice Brandeis famously propounded and as courts and commentators often suggest? If so, why? Why are these intellectual goods deemed the "noblest"? Why should they necessarily be placed in the commons? The exclusion of (abstract) ideas from patent and ideas from copyright is a foundational and yet somewhat confusing area of intellectual property law.

Despite the bedrock nature of the general rule that the fundamental intellectual building blocks cannot and should not be subject to the embarrassment of exclusive rights,[91] there has not been much attention devoted to understanding *why* this should be the case. There have been numerous pronouncements by courts declaring ideas to be outside the intellectual property system, and the copyright statute expressly excludes ideas and other intellectual building blocks. Often the rhetoric employed by courts has analogized ideas to fundamental forces or products of nature.[92] As Jefferson famously suggested:

> 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 in any point, and like the air in which we breathe, move, and have our physical being, incapable of confinement or exclusive appropriation.[93]

[91] Jefferson (1813).

[92] O'Reilly v. Morse, 15 How. 62 (1854)

[93] Jefferson (1813).

Ideas are not forces or products of Nature. Ideas are intangible products of the human intellect. As such, why shouldn't they be subject to exclusive property? The short answer is that ideas are often, though not always, intellectual infrastructure. Jefferson recognized the special infrastructural characteristics of ideas; the quote above indicates that he not only took notice of the nonrivalrous nature of ideas, but also saw how ideas are productive inputs, "for the moral and mutual instruction of man, and improvement of his condition."[94] The noble stature of ideas seems to be rooted in appreciation of both the instrumental nature of ideas as means to human flourishing and societal progress, and the egalitarian potential of leveraging nonrivalry so that everyone can benefit and have the capacity to participate.[95]

This section is divided into two subsections. The first discusses infrastructural ideas, and the second examines legal recognition of the infrastructural nature of ideas and the social value of commons management, focusing on the First Amendment, copyright law, and patent law. I argue that despite some confusion and controversy on how to draw lines and separate ideas from expression and invention (which I suggest are outputs from particular uses of ideas), ideas are and should be "free as the air to common use."[96]

### a. Ideas as Infrastructure

Ideas are a particularly good example of intellectual infrastructure, because they are nonrival inputs into a wide variety of productive uses.[97] Broadly categorized, the scope of uses includes (1) further idea production/research, (2) expression/communication/education, and (3) implementation/transformation of the physical or social world. Many others have explored this metaphysical terrain in a more sophisticated and detailed fashion.[98] My point here is to segregate uses of ideas roughly along functional lines. Frankly, the lines also coincide with legal categories, as we will see below.

Ideas implicate each of the three boundary issues discussed in the previous section: First, it is difficult to draw lines where there is a stream of cumulative inputs, as is often the case with ideas. Second, it is difficult to choose the appropriate level of abstraction. Third, it is difficult to choose the appropriate scope of uses. The line-drawing consideration

---

[94] *Id.* Jefferson was not the first to recognize the power of ideas. His views reflect a much broader historical tradition.

[95] We might say that ideas are "essential" and "affected with the public interest" in a similar way as traditional infrastructure. See chapter 5.

[96] International News Serv. v. Associated Press, 248 U.S. 215, 250 (1918) (Brandeis, J., dissenting).

[97] Frischmann & Lemley 281 (2007) ("intangible infrastructure, such as ideas, . . . may be the cleanest example of the benefits of commons because the advantage of private ownership in solving the tragedy of the commons does not apply to information, which is inexhaustible. Ideas themselves are a good example of infrastructure, because they are not merely passively consumed but frequently are reused for productive purposes.") (footnote omitted).

[98] See, e.g., Locke (1975); Hume (1986).

arises directly and acutely in the context of intellectual property. For present purposes, I leave it aside and focus mostly on the latter two difficulties.

Ideas constitute intellectual infrastructure when we consider ideas on an abstract, philosophical level. Even at less abstract levels, ideas are infrastructural. So when are ideas not infrastructural? Some ideas are inputs into a rather limited range of uses or outputs. In the previous section, I discussed two examples of ideas that were not infrastructural—the idea of a common construction wire nail and the idea of the 386 chip. Both ideas are inputs into a rather narrow range of outputs—the common nail and the 386 chip. Recall that this discussion raised boundary issues because applying the infrastructure criteria depended on the level of abstraction. For example, I concluded that the relatively concrete (applied) idea of the 386 chip would not qualify as infrastructure, but the more abstract (basic) idea of backward compatibility would. Similarly, I suggested that the idea of a common nail is not infrastructural because the idea is an input into the production of a tangible nail and not much else. Again, the idea is rather concrete (applied) and tied to the specific tangible embodiment. To expand the range of outputs rather dramatically, we could drop the specifications of common construction wire nail. At a more abstract level, the idea of a nail—expressed as, the idea of a pin-shaped fastener or the idea of a fastener that holds materials together by shear strength laterally and friction axially—is infrastructural. The more abstract idea of a nail leaves considerable room for variety in implementation, for example, in terms of design, the material or process used to produce the nail, the dimensions, and so on.

Thus far, I have narrowed the *scope* of relevant uses of the idea by focusing on implementation of the idea through transformation of physical resources (i.e., producing tangible embodiments). The reason for doing so is that these uses are most prevalent from the demand side. I could have mentioned uses that involve further idea production, such as research and development. Also, I could have mentioned uses that involve communicating the idea, such as expression—the idea of a nail is certainly used expressively in instructions, plans, illustrations, or even metaphorically (e.g., the gymnast "nailed" the landing). Expanding the scope of relevant uses adds variety and thus might lead one to more easily classify the idea as infrastructural. In fact, this expansion would apply to the abstract idea as well as the more concrete ideas of a common nail and the 386 chip. Admittedly, I hesitate to include such additional uses in all cases, for the same reason that when discussing the rivalrousness of an apple in chapter 3, I narrowed the scope of relevant uses by excluding some possible nonrivalrous uses of an apple, such as using an apple as the subject matter of a painting or photograph. Possible does not mean relevant. The uses seem considerably less important from the demand side. The vast majority of social demand for (value attributable to) apples is derived from rivalrous consumption. Similarly, the vast majority of social demand for (value attributable to) the idea of a common nail is derived from the production of tangible nails.

Yet this conclusion depends substantially on what level of abstraction we are working at. The more abstract idea of a nail may yield significant social value both from (a) further

idea production, research and development, or simply refinement of the abstract idea to make it more concrete and capable of implementation, and (b) expressive use that communicates the idea so people may internalize it as knowledge. This example demonstrates another way in which ideas are quite different from apples. It also shows why it makes sense to categorize ideas as intellectual infrastructure, at least as a default.

b. Commons Management via First Amendment, Copyright, and Patent Jurisprudence

The infrastructural nature of ideas and the social value of sustaining the public commons appear to be reasonably well established in various areas of the law, especially the First Amendment and intellectual property. The First Amendment is fundamentally about ideas. It protects the freedoms of speech and of the press from government interference.[99] As chapter 3 discussed, speech involves the communication of ideas. The same can be said of the press.[100] In the First Amendment context, ideas are recognized as a basic input for a host of socially valuable activities, including public debate, discourse, and education on commercial, political, and various other societal issues. Consider political speech, for example.

> Political speech, at the core of protected First Amendment speech, involves the communication of ideas used productively in political systems. The public good nature of ideas enables repeated sharing and productive use, which often has dynamic and systemic implications in political systems that speakers may not anticipate or appreciate fully. As Posner suggests, because political speech generates many different types of spillovers the category is especially susceptible to underproduction. Government regulation, which Posner helpfully analogizes to a tax, would only exacerbate the problem. As important as the problem of underproduction is the serious concern about error costs from misdirected regulation. Posner, Farber, and many others have noted that First Amendment protection is especially needed for political speech because otherwise political speakers (candidates, political parties, etc.) might introduce considerable bias into the system. The First Amendment constraint on government intervention sustains the flow of spillovers from the repeated sharing and productive use of ideas communicated through political speech.[101] In other words, the First Amendment recognizes the complex relationships

[99] "Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances." Though tempting, I leave aside discussion of religious freedom and freedom of association, both of which support the basic argument I am making.

[100] Baker (2007).

[101] This paragraph draws from Frischmann 323–24 (2008b). See also Posner 22–23 (1986); Farber 563 (1991); Farber 935 (2006).

> between the activity of speech, the power of communicated ideas, and the consequences for political and other social systems. Nothing less than the fate of democracy itself is said to rest on our First Amendment freedoms and the various idea-dependent activities and social practices those freedoms support.

Freedom of speech is frequently associated with the metaphor of a "marketplace of ideas." Justice Holmes is credited with bringing the concept into the First Amendment jurisprudence when he stated that it may be logical for people who hold strong opinions about the truth or superiority of their ideas to try to fix those ideas "in law and sweep away all opposition," but "the ultimate good desired is better reached by free trade in ideas . . . that the best test of truth is the power of the thought to get itself accepted in the competition of the market, and that truth is the only ground upon which their wishes safely can be carried out."[102] As many have pointed out, sustaining an open and vibrant "marketplace of ideas" is *not* about sustaining an actual market with transactions mediated by the price mechanism; rather, the marketplace of ideas metaphor concerns an open exchange of ideas, unbiased competition among many different ideas, and a diverse and wide range of competitors (idea producers and users).[103] In essence, the "marketplace of ideas" metaphor reflects a strong commitment to a public commons.

The First Amendment not only constrains the government's ability to "pick winners" directly, for example through censorship or favoring particular viewpoints, but also constrains the government's ability to eliminate the public commons.[104] At a macro level, at least, the First Amendment limits the extent to which the government can privatize and propertize ideas. In *Eldred v. Ashcroft*, the Supreme Court described the exclusion of ideas from copyright protection as a "built-in First Amendment accommodation" that "strike[s] a definitional balance between the First Amendment and the Copyright Act by permitting free communication of facts while still protecting an author's expression." The Court explained that "[d]ue to this distinction, every idea, theory, and fact in a copyrighted work becomes instantly available for public exploitation at the moment of

---

[102] Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes dissenting). Holmes recognized that ideas are means to "the ultimate good desired."

[103] For further discussion, see Frischmann n.6 (2008b); Netanel 158–60 (2005).

[104] The First Amendment plays other important roles in structuring the cultural environment. See Balkin (2009); Balkin (2004). For example, the First Amendment may impose an obligation on government to ensure sufficient speech capabilities for its citizens. Such capabilities may depend on various shared public infrastructures (e.g., public spaces, communications technologies, laws). See *id.* (advancing the concept of "an infrastructure of free expression"); Ammori (forthcoming 2012) (arguing for the First Amendment requires government provision, or at least assurance of availability, of sufficient public "spaces"). I agree with much of what Balkin and Ammori have to say, but leave aside this topic for future research.

publication."[105] The First Amendment can be understood as a constitutional commitment to public commons in ideas.[106]

Similarly, based on a firm recognition that ideas are basic inputs for cultural, scientific, and technological progress, the copyright and patent laws exclude ideas from protection and allocate ideas to the public commons, generally referred to as the public domain.[107] In theory and formally, copyright law extends protection only to original expression of ideas,[108] and patent law extends protection only to new, useful, and nonobvious inventions,[109] which constitute implementations or embodiments of ideas. Put another way, the intellectual property system relies on different bodies of law to address different subsets of idea uses—patent law addresses uses related to physical implementation, and copyright law addresses uses related to expression. Patent law is concerned with whether the idea can be implemented; if so, then a patent may issue. Copyright law, on the other hand, attaches to particular expressions of the idea, for example an illustration or textual description. Patent law does not aim to regulate expression; to the extent that patent law deals with expression, it generally pushes expression of ideas toward the public commons (e.g., via public disclosure, written description) to ensure public dissemination of the expressed ideas. Similarly, copyright law does not aim to regulate idea implementation; to the extent that copyright law encounters issues pertaining to implementation, it generally pushes implementations or embodiments of ideas toward patent law or the public commons. Neither patent nor copyright directly addresses the use of ideas to generate more ideas. Ideas generated in cumulative stages of idea production, research and development, and so on remain outside the scope of patent and copyright. Copyright and patent laws grant exclusive rights only to certain qualifying outputs produced from the use of ideas, leaving ideas "free as the air to common use," at least after voluntary communication to others.[110]

In reality, the intellectual property regimes struggle mightily to separate unprotectable ideas from copyrightable expression and patentable invention. The struggle stems in part

---

[105] Eldred v. Ashcroft, 537 US 186, 219 (2003) (citing Feist, 499 U.S., at 349–350). See also Harper & Row, Publrs. v. Nation Enters., 471 U.S. 539, 556 (1985). Of course, it also may be understood to reflect other commitments. See previous note. There is a rich literature on the relationship between the First Amendment and copyright law. See, e.g., Netanel (2008); Benkler (1999); Baker (1997); Lessig (2001a); Patterson (1987); Nimmer (1970). On how the First Amendment affects congressional power in patent, see Burk (2000); Sawkar 3048 (2008); Pollack (2002)

[106] Frischmann (2008b).

[107] 17 U.S.C. § 102(b) ("In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery . . ."); Diamond v. Chakrabarty, 447 U.S. 303, 309 (1980) ("The laws of nature, physical phenomena, and abstract ideas have been held not patentable.").

[108] 17 U.S.C. § 102 (a) & (b).

[109] 35 USC 101, 102, 103.

[110] INS. v. AP (Brandeis, J., dissenting).

from many of the complications noted in the previous section—for example, the dynamic and complex nature of intellectual systems, the continuity or fluidity of resource streams, and dual product-process nature of intellectual resources. Typically, ideas are not discrete things that can be easily identified, delineated, and isolated from other ideas or the context within which they are developed and their meaning derives and evolves. An idea can be stated or implemented; such actions effectively construct a discrete thing bounded by the specificity of action and the output (i.e., the statement or implementation), but the output/thing is not the idea.

Expression and invention describe both the action and the output. That is, "expression" ("invention") describes the act of stating (implementing) and the statement (implementation). While this duality of meaning may pose some complications, expression and invention are discrete things that can be identified, delineated, and isolated, at least when manifest in tangible form such as a written note or a working embodiment.

If copyright and patent protection only extended to such discrete things—that is, if copyright only protected the literal expression fixed in the tangible form of the note and patent only protected the actual working embodiment created by the inventor and disclosed in a patent application—then the struggle to separate unprotectable ideas from copyrightable expression and patentable invention would probably go away. But copyright protection and patent protection are not so limited in scope, nor should they be. In both areas, courts have explained why the scope of protection must be broader if the legal regimes are to effectively serve their purpose of promoting Progress in Science and the useful Arts:[111]

- "It is of course essential to any protection of literary property . . . that the right cannot be limited literally to the text, else a plagiarist would escape by immaterial variations. That has never been the law, but, as soon as literal appropriation ceases to be the test, the whole matter is necessarily at large, so that . . . the decisions cannot help much in a new case. . . ."[112] In essence, the court suggests that once the scope of protection exceeds the literal text, each case must be decided carefully on its own facts.
- "Courts have also recognized that to permit imitation of a patented invention which does not copy every literal detail would be to convert the protection of the patent grant into a hollow and useless thing. Such a limitation would leave room for—indeed, encourage—the unscrupulous copyist to make unimportant and insubstantial changes and substitutions in the patent which, though adding nothing, would be enough to take the copied matter outside the claim, and hence outside the reach of law. One who seeks to pirate an invention, like one who

[111] US CONST. art. I, § 8, cl. 8.

[112] Nichols v. Universal Pictures Corp., 45 F. 2d 119 (2d Cir. 1930).

> seeks to pirate a copyrighted book or play, may be expected to introduce minor variations to conceal and shelter the piracy. Outright and forthright duplication is a dull and very rare type of infringement."[113]

The recognized need to expand scope to preclude "plagiarism by immaterial variation" and "copying by unimportant and insubstantial changes and substitutions" resonates with classic misappropriation. It reflects, in economic terms, the basic supply-side problems described earlier in the chapter: high exclusion costs and the associated risk that competitors will free ride on the fixed cost investment of the author or inventor. Note that the recognized need to expand the scope of protection to exclude this type of free riding does not resolve a series of fundamental questions: How *much* should scope be expanded? In what ways should scope be expanded? Recall how the scope of exclusion can vary in many different dimensions.

The struggle to separate unprotectable ideas from copyrightable expression and patentable invention thus stems from the difficulty in choosing the appropriate scope of protection and, consequently, in choosing the appropriate "level of abstraction" to effectuate that choice.[114] The *scope* of relevant uses is more or less chosen by the institutional structure, with patent focused on implementing uses, and copyright focused on expressive/communicative uses. I briefly explain how this abstraction issue arises in both copyright and patent law.

In copyright law, Learned Hand famously introduced a level of abstractions analysis:

> Upon any work, and especially upon a play, a great number of patterns of increasing generality will fit equally well, as more and more of the incident is left out. The last may perhaps be no more than the most general statement of what the play is about, and at times might consist only of its title; but there is a point in this series of abstractions where they are no longer protected, since otherwise the playwright could prevent the use of his "ideas," to which, apart from their expression, his property is never extended. . . . Nobody has ever been able to fix that boundary, and nobody ever can.[115]

Despite his acknowledgment of the inherent struggle to draw a clear line between idea and expression, Hand proceeded to evaluate the similarities and differences between two

---

[113] Graver Tank & Mfg. Co. v. Linde Air Prods. Co., 339 U.S. 605, 607 (1950) ("The doctrine of equivalents evolved in response to this experience. . . . The theory on which it is founded is that, 'if two devices do the same work in substantially the same way, and accomplish substantially the same result, they are the same, even though they differ in name, form or shape.'") (quoting Union Paper-Bag Machine Co. v. Murphy, 97 U. S. 120, 125 (1877)).

[114] Chiang (2011).

[115] Nichols, 45 F. 2d at 122.

dramatic works (a play and a motion picture) at different levels of abstraction, focusing primarily on the "the characters and sequence of incident." The following is illustrative of the approach:

> If *Twelfth Night* were copyrighted, it is quite possible that a second comer might so closely imitate Sir Toby Belch or Malvolio as to infringe, but it would not be enough that for one of his characters he cast a riotous knight who kept wassail to the discomfort of the household, or a vain and foppish steward who became amorous of his mistress. These would be no more than Shakespeare's "ideas" in the play, as little capable of monopoly as Einstein's Doctrine of Relativity, or Darwin's theory of the Origin of Species. It follows that the less developed the characters, the less they can be copyrighted; that is the penalty an author must bear for marking them too indistinctly.
>
> In the two plays at bar we think both as to incident and character, the defendant took no more . . . than the law allowed. The stories are quite different. One is of a religious zealot who insists upon his child's marrying no one outside his faith; opposed by another who is in this respect just like him, and is his foil. Their difference in race is merely an obbligato to the main theme, religion. They sink their differences through grandparental pride and affection. In the other, zealotry is wholly absent; religion does not even appear. It is true that the parents are hostile to each other in part because they differ in race; but the marriage of their son to a Jew does not apparently offend the Irish family at all, and it exacerbates the existing animosity of the Jew, principally because he has become rich, when he learns it. . . . The only matter common to the two is a quarrel between a Jewish and an Irish father, the marriage of their children, the birth of grandchildren and a reconciliation.
>
> If the defendant took so much from the plaintiff, it may well have been because her amazing success seemed to prove that this was a subject of enduring popularity. . . . Though the plaintiff discovered the vein, she could not keep it to herself; so defined, the theme was too generalized an abstraction from what she wrote. It was only a part of her "ideas."[116]

In this framework, ideas are often assumed to be abstract, basic, or generalized inputs in the sense that ideas can be expressed in many different ways. This conforms to the basic rule that ideas are unprotected and expression is protected by copyright. Yet in applying the framework, courts evaluate whether ideas at various levels of abstraction are being expressed in a work and whether the scope of copyright protection extends to such expression. At intermediate levels of abstraction, ideas may be used communicatively (i.e., to generate expression), or may be used to produce more refined ideas at lower levels

[116] *Id.*

of expression (for example, an idea of a love story may be used to generate the more refined idea of a love story plot involving people from different religious backgrounds, which might in turn be used to generate an even more refined idea of a love story plot, etc.). In addition, for certain types of works, such as software, ideas at various levels of abstraction implement functional rather than expressive ends.[117] Courts have developed more sophisticated approaches to the levels-of-abstraction analysis. For example in *Computer Associates v. Altai*, the Court of Appeals for the Second Circuit developed an Abstraction-Filtration-Comparison approach with which a court filters out a range of unprotected elements at various levels of abstraction from the copyright owner's software program before comparing it with the defendant's program.[118]

Moreover, in some cases, courts recognize the existence of ideas that can only be expressed effectively in a limited number of ways, though the idea may be useful in many other ways and thus still infrastructural. For example, the idea underlying an algorithm may be expressed in a mathematical equation, and there may be a limited number of ways to effectively express the idea. Courts have developed a "merger doctrine" in such cases to limit protection.[119] In developing the doctrine, courts were concerned with copyright protection of the expression leading to de facto protection of the idea. The limitation imposed by this doctrine varies among different Courts of Appeals: In some jurisdictions no copyright protection is permitted, and in others protection is deemed "thin," meaning that the scope of protection is limited to literal or identical copying.[120]

The levels-of-abstraction framework has become a regular feature of copyright law, as a methodology employed in infringement analysis to separate unprotected ideas from protected

---

[117] Computer Associates International, Inc. v. Altai, Inc., 982 F. 2d 693, 712 (2d Cir. 1992); Lotus Development Corporation v. Borland International, Inc., 49 F. 3d 807, 815 (1st Cir. 1995). Judge Boudin's concurring opinion in Borland is particularly attuned to the additional social costs of limiting access to infrastructural elements of software programs and interfaces:

> Of course, the argument for protection is undiminished, perhaps even enhanced, by utility: if we want more of an intellectual product, a temporary monopoly for the creator provides incentives for others to create other, different items in this class. *But the "cost" side of the equation may be different where one places a very high value on public access to a useful innovation. . . . Thus, the argument for extending protection may be the same; but the stakes on the other side are much higher.*

Lotus Dev. Corp. v. Borland Int'l, Inc., 49 F. 3d 807, 819 (1st Cir. 1995) (Boudin, J., concurring) (emphasis added).

[118] Altai, 982 F. 2d 693, at 706.

[119] Toro Company v. R&R Products Co., 787 F. 2d 1208, 1212 (8th Cir. 1986) ("Under the copyright law doctrine of merger, a close cousin to the idea/expression dichotomy, copyright protection will be denied to even some expressions of ideas if the idea behind the expression is such that it can be expressed only in a very limited number of ways. The doctrine is designed to prevent an author from monopolizing an idea merely by copyrighting a few expressions of it."); Morrissey v. Procter & Gamble Co., 379 F. 2d 675, 678–79 (1st Cir. 1967) (denying copyright protection to the wording of rules for a sweepstakes contest, on account of the limited number of ways to express the rule).

[120] *Id.*; Apple Computer Corp. v. Microsoft Corp., 33 F. 3d 1435, 1439–42 (9th Cir. 1994) (thin protection); see also Nimmer on Copyright 13.03[B][3] (2010).

expression and more broadly as a conceptual tool for understanding the scope of the copyright system and its relationship with the First Amendment. The framework transforms the bright-line rule of "ideas out, expression in" into a context-specific standard to be applied on a case-by-case basis. There is no bright line rule telling courts (or anyone else) how to determine the "optimal scope" of copyright protection or how to choose the level of abstraction,[121] perhaps because such a rule cannot exist.[122] As a result, courts (and everyone else) must muddle ahead with an understanding of the struggle, the need to evaluate and determine scope in context, and healthy appreciation of the default rule (ideas out, expression in).

In patent law, courts also struggle with separating unprotectable ideas from patentable invention. The struggle is complicated by the fact that patent law is not very clear about the baseline rule of excluding ideas. In contrast with copyright, the patent statute does not refer to "ideas," much less expressly exclude them from protection. Instead, the patent statute focuses on "invention."[123]

Courts and commentators are all over the map on what constitutes an invention—is it an idea, or is it the implementation or tangible embodiment of an idea? Tun-Jen Chiang aptly describes the definitional confusion:

> When forced to clarify what an "invention" really is, however, leading authorities take directly contradictory approaches. . . . Chief Judge Howard Markey of the Federal Circuit, one of the preeminent judges of patent law, has characterized an

[121] This has led to conflicting views on the merits of the approach. Compare, for example, Chiang [draft at 50] (2011) ("the abstractions test [] provide[s] an enormously useful framework, reminding judges of 'the difficulties that require courts to avoid either extreme of the continuum of generality.'") (quoting Nash v. CBS, Inc., 899 F. 2d 1537, 1540 (7th Cir. 1990)) with Cohen 732 (1987) (criticizing the doctrine for leading to "unpredictable, impressionistic" decisions) and Yen (2003) (criticizing idea-expression and fair use doctrines resort to abstraction and noting that "[i]t is dangerous to put free speech at the mercy of the idea/expression dichotomy and fair use because those doctrines do not have enough substance to adequately protect something so important."). Most recognize that the approach is imperfect because it is, among other things, inherently messy, unpredictable, and subjective. Nonetheless, it may be the best available means for evaluating and tailoring the scope of protection to the intellectual work.

[122] Cf. Fromer 745 (2009) ("Fixing the boundary between idea and expression can be difficult, not only because of the line drawing required to determine which abstractions of the expression are still protected enough to be more of an expression than an idea, but also because there is no sharp ex ante sense of what the copyright protects beyond the copyrighted work itself").

[123] The statute provides: "Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title." 35 USC 101. The Supreme Court broadly construed "process, machine, manufacture, or composition of matter," noting that Congress chose expansive terms and even modified them with the "comprehensive 'any.'" Diamond v. Chakrabarty, 447 U.S. 303 (1980). In Chakrabarty, the Court concluded, "Congress intended statutory subject matter to 'include anything under the sun that is made by man.'" *Id.* (quoting S. Rep. No. 1979, 82d Cong., 2d Sess., 5 (1952); H. R. Rep. No. 1923, 82d Cong., 2d Sess., 6 (1952)). Immediately following this expansive construction of patentable subject matter, the Court noted, "This is not to suggest that 101 has no limits or that it embraces every discovery. The laws of nature, physical phenomena, and abstract ideas have been held not patentable." *Id.*

"invention" as the embodiment described by a specification. According to Chief Judge Markey, "Ideas are never patentable. Only embodiments of an idea, i.e. an invention, may be patented." Moreover, "idea" is a "mud word" that "appears nowhere in the statute, which speaks only of invention." This view of invention as embodiment is supported by some Supreme Court precedent as well as the statute that defines patentable inventions as tangible machines, products, and processes that can be made and used.[124]

On the other side, Judge Giles Rich of the Federal Circuit, another eminent judge who wrote much of the 1952 Patent Act, has opined that an "invention" is an abstract idea. According to Judge Rich, an invention is "an incorporeal, intangible abstraction in the nature of a product of the mind." An embodiment is "[p]opularly but inaccurately called 'invention.'" This view, too, has support in Supreme Court precedent, which states that "[t]he primary meaning of the word 'invention' in the Patent Act unquestionably refers to the inventor's conception rather than to a physical embodiment of that idea."[125]

On top of this confusion over what constitutes an invention, the Supreme Court has long held that "laws of nature, physical phenomena, and *abstract ideas*" are not patentable subject matter and that "[t]he concepts covered by these exceptions are 'part of the storehouse of knowledge of all men . . . free to all men and reserved exclusively to none.'"[126] Obviously, this rule mirrors the idea/expression dichotomy in copyright law.[127] However, it still leaves unclear what constitutes an abstract idea and what constitutes an invention.

*Abstract ideas excluded*

In *Bilski v. Kappos* (2010), the Supreme Court reiterated the rule that abstract ideas are not patentable.[128] The case involved a person trying to patent a method for hedging risk in energy markets. According to the Court, the key claims were 1 and 4; claim 1 described

[124] Chiang (2011) (citing Markey 333 (1983); Seymour v. Osborne, 78 U.S. 516, 552 (1870); 35 U.S.C. § 101); Rubber-Tip Pencil Company v. Howard, 87 U.S. (20 Wall.) 498, 507 (1874) ("An idea of itself is not patentable, but a new device by which it may be made practically useful is.").

[125] Chiang (2011) (citing Rich (1942)); Pfaff v. Wells Elecs., Inc., 525 U.S. 55, 60 (1998); Gill v. United States, 160 U.S. 426, 434 (1896) ("In every case the idea conceived is the invention.")).

[126] Bilski v. Kappos, 130 S.Ct. 3218, 561 U.S. _ (2010) (quoting Funk Brothers Seed Co. v. Kalo Inoculant Co., 333 U. S. 127, 130 (1948)). In many of the cases reiterating that basic rule, the Court has stated more broadly that ideas as such are not patentable. For example, in Benson, the Supreme Court stated, "It is conceded that one may not patent an idea." 409 U.S., at 71.

[127] Landes & Posner 305 (2003).

[128] The rule is rooted in both English and American law. See, e.g., Neilson v. Harford, Webster's Patent Cases 295, 371 (1841); Le Roy v. Tatham, 14 How. 156. 175 (1853); O'Reilly v. Morse, 15 How. 62 (1854); The Telephone Cases, 126 U. S. 1 (1888).

a series of steps instructing how to hedge risk, and claim 4 provided a formula that restated in mathematical terms the concept in claim 1. The Court "resolve[d] th[e] case narrowly on the basis of [its] decisions in *Benson*, *Flook*, and *Diehr*, which show that [Bilski's] claims are not patentable processes because they are attempts to patent abstract ideas. Indeed, all members of the Court agreed that the patent application at issue fell outside of §101 because it claims an abstract idea."[129] The Court discussed the three decisions and applied the underlying lessons as follows:

- In *Benson*, the Court considered whether a patent application for an algorithm to convert binary-coded decimal numerals into pure binary code was a "process" under §101.[130] The Court first explained that "'[a] principle, in the abstract, is a fundamental truth; an original cause; a motive; these cannot be patented, as no one can claim in either of them an exclusive right.'"[131] The Court then held the application at issue was not a "process," but an unpatentable abstract idea. "It is conceded that one may not patent an idea. But in practical effect that would be the result if the formula for converting . . . numerals to pure binary numerals were patented in this case."[132] A contrary holding "would wholly pre-empt the mathematical formula and in practical effect would be a patent on the algorithm itself."[133]
- In *Flook*, the Court considered the next logical step after *Benson*. The applicant there attempted to patent a procedure for monitoring the conditions during the catalytic conversion process in the petrochemical and oil-refining industries. The application's only innovation was reliance on a mathematical algorithm.[134] *Flook* held the invention was not a patentable "process." The Court conceded the invention at issue, unlike the algorithm in *Benson*, had been limited so that it could still be freely used outside the petrochemical and oil-refining industries.[135] Nevertheless, *Flook* rejected "[t]he notion that post-solution activity, no matter how conventional or obvious in itself, can transform an unpatentable principle into a patentable process."[136] The Court concluded that the process at issue there was "unpatentable under §101, not because it contain[ed] a mathematical algorithm as one component, but because once that algorithm [wa]s assumed to be within the prior art, the application, considered as a whole, contain[ed] no

[129] Bilski, 130 S.Ct. at 3230.

[130] 409 U.S., at 64–67.

[131] *Id.* at 67 (quoting Le Roy, 14 How., at 175).

[132] 409 U.S., at 71.

[133] *Id.* at 72.

[134] 437 U.S. at 585–586.

[135] 437 U.S. at 589–590.

[136] *Id.* at 590.

patentable invention."[137] As the Court later explained, *Flook* stands for the proposition that the prohibition against patenting abstract ideas "cannot be circumvented by attempting to limit the use of the formula to a particular technological environment" or adding "insignificant postsolution activity."[138]

- Finally, in *Diehr*, the Court established a limitation on the principles articulated in *Benson* and *Flook*. The application in *Diehr* claimed a previously unknown method for "molding raw, uncured synthetic rubber into cured precision products," using a mathematical formula to complete some of its several steps by way of a computer.[139] *Diehr* explained that while an abstract idea, law of nature, or mathematical formula could not be patented, "an application of a law of nature or mathematical formula to a known structure or process may well be deserving of patent protection."[140] *Diehr* emphasized the need to consider the invention as a whole, rather than "dissect[ing] the claims into old and new elements and then . . . ignor[ing] the presence of the old elements in the analysis."[141] Finally, the Court concluded that because the claim was not "an attempt to patent a mathematical formula, but rather [was] an industrial process for the molding of rubber products," it fell within §101's patentable subject matter.[142]
- In light of these precedents, it is clear that [Bilski's] application is not a patentable "process." Claims 1 and 4 in petitioners' application explain the basic concept of hedging, or protecting against risk: "Hedging is a fundamental economic practice long prevalent in our system of commerce and taught in any introductory finance class."[143] The concept of hedging, described in claim 1 and reduced to a mathematical formula in claim 4, is an unpatentable abstract idea, just like the algorithms at issue in *Benson* and *Flook*. Allowing petitioners to patent risk hedging would preempt use of this approach in all fields, and would effectively grant a monopoly over an abstract idea.

The Court's "analysis," in the excerpts quoted above, is not particularly helpful as guidance for differentiating abstract ideas from patentable inventions.[144] The Court does not indicate any reasons or characteristics that distinguish abstract ideas from nonabstract ideas. As Justice Stevens noted in his concurrence, "The Court, in sum, never provides a

[137] *Id.* at 594.

[138] Diehr, 450 U.S., at 191–92.

[139] 450 U.S. at 177.

[140] *Id.* at 187.

[141] *Id.* at 188.

[142] *Id.* at 192–93.

[143] 545 F. 3d at 1013 (Rader, J., dissenting).

[144] See Stevens concurrence; Schultz & Samuelson (2011).

satisfying account of what constitutes an unpatentable abstract idea."[145] Rather, the Court bases its conclusion fully on the consequences of classification for patent scope; that is, the concept of hedging and the mathematical formula are abstract ideas, says the Court, because of the preemptive effect of deciding otherwise.

A better way to understand the decision and the rule it was attempting to describe and apply is to view the analysis as identical to the copyright inquiry: Bilski's invention cannot be extended to the level of abstraction reflected in claims 1 and 4 because that would allow the scope of the patent to reach beyond the embodiment of the idea disclosed in the patent application and the family of related embodiments that reflect Bilski's inventive contribution. To make more sense of this alternative view, I turn to the question of what constitutes an invention.

*Invention and patent scope*

Chiang argues persuasively that invention has two different meanings in two separate but related contexts in patent law: First, invention refers to the embodiment, the "tangible and working apparatus or process" described in the patent specification.[146] Using an idea to create such a new and useful "process, machine, manufacture, or composition of matter" is what entitles a person to a patent. Second, invention refers to the claimed idea, rather than the specific embodiment. In the infringement context, courts look to the claims to define the scope of the patent, and in this context, invention takes on a broader meaning than the embodiment, for the reason noted earlier. Chiang argues that courts determine patent scope by choosing the level of abstraction with which to analyze infringement, and that they do so "*implicitly* and on an ad hoc basis." He suggests that patent law should follow copyright law and acknowledge the levels-of-abstraction problem, its necessity in the context of differentiating unprotected ideas and protected inventions, and the policy discretion involved in the process. I largely agree with his analysis. However, I would push toward a more explicit exclusion of ideas from patent and frame the analysis in a slightly different manner, described in more detail in the following paragraphs.

As noted, invention describes the act of implementing and the implementation. In contrast with ideas, an invention is a discrete thing that can be identified, delineated, and isolated. *Invention is the working embodiment described by a specification.* I agree with Judge Markey and the associated line of cases and commentary.

For patent rights to be *effective as means*, the rights cannot be limited in scope to the disclosed embodiment but instead must be extended to cover a family of related embodiments. The function of claims is to delineate the family of related embodiments.

---

[145] Stevens concurrence (He goes on: "The Court essentially asserts its conclusion that petitioners' application claims an abstract idea").

[146] Chiang (2011).

- Important tangent: *Effectiveness as means* depends, in turn, on what is the relevant economic objective or end. Despite plenty of rhetoric and wrangling, the basic question of whether patent law aims (a) to facilitate markets and correct for supply-side risks associated with high exclusion costs and the cost structure of supply—to facilitate average cost recovery and no more—or (b) to put an additional weight on the incentives to invest scale remains unsettled. The fact that patents provide the power to exclude independent invention seems to push toward the latter objective,[147] but the issue is still open and contentious.[148]

Each embodiment is actually the product of many different ideas at various levels of abstraction, most of which are not novel. Thus, claims should be understood to identify the shared *idea* among the family of embodiments covered by the patent that constitutes the patentee's contribution *to the public domain*. This is where my framing differs from Chiang's. He suggests that the claimed idea is within the scope of the patent and thus owned. I argue that the idea remains in the public domain, but it serves as an indicator for the family of patented embodiments.[149] The public remains free to use the idea, for example, to communicate knowledge or to generate further ideas (e.g., research, workarounds, or improvements). The patent only precludes making, using, and selling embodiments in the family.[150]

Interpreting, construing, and applying patent claims is fundamentally the same exercise that courts perform in copyright cases when they engage in levels-of-abstraction analysis or otherwise attempt to separate ideas from expression. As Chiang argues, claims construction involves evaluating and choosing an appropriate level of abstraction for the idea shared among protected embodiments and consequently the scope of patent protection.[151]

---

[147] According to Cotropia & Lemley (2009), most patent litigation does not involve copying.

[148] See, e.g., Maurer & Scotchmer (2002); Vermont (2006); Lemley (2007). Even if the economic objective is to spur investment in inventive activities above and beyond what average cost recovery would support, it remains unclear that including independent invention within the scope of the patent (rather than recognizing an independent invention defense) is the most efficient way to accomplish that objective. It may very well be the case that other mechanisms for expanding patent scope would be more efficient.

[149] Similarly, Jeanne Fromer describes patent claims as "listing [the] necessary and sufficient characteristics" that delineate "the set of protected embodiments." Fromer 721 (2009). Fromer provides a detailed examination of claiming in copyright and patent law and the scope issues associated with different types of claiming rules. Beyond the fundamental scope issue, she also discusses various costs and benefits of different claiming rules. *Id.* at 722.

[150] *Id.* at 721.

[151] Courts routinely engage is this type of analysis when construing elements of a patent claim and determining the scope of protection afforded by patent. For example, as Burk and Lemley observe:

> In determining the meaning of terms within a particular element, judges practicing patent claim interpretation are engaged in an exercise that to some degree resembles the famous "levels of abstraction test" articulated by Judge Learned Hand for analysis of infringement under copyright law's "idea/expression"

Determining what embodiments populate the set depends on abstraction analysis. The set of embodiments that share an idea may be infinite and growing over time; not all of its members may exist or be enabled at the time of patent filing. The same issue arises in the context of the doctrine of equivalents when courts determine whether embodiments that are not literally within the scope of the claims nonetheless are within the scope of patent protection.

Patent scope is and must be determined through abstractions analysis. Acknowledging the importance of abstraction leaves the door open for tailoring patent scope to the context and type of invention.[152] In some contexts, a low level of abstraction might be chosen to grant "thin" protection, effectively limiting the patent scope to a narrow range of embodiments, perhaps only those embodiments that are currently enabled. In other contexts, a somewhat higher level of abstraction might be chosen to confer "thick" protection, extending protection to embodiments that do not currently exist (or are not currently enabled) but are on the foreseeable horizon. Patent scope can be extended even further, for example, to embodiments that are not foreseeable. I do not make a strong claim about what the scope of patent protection ought to be in general, much less in specific contexts, because such a policy determination depends on the economic and other social objectives, and, as noted, it is not entirely clear that we know what the economic objective is for patent law.

- To illustrate how the underlying economic objective matters, consider the doctrine of equivalents (DOE). The DOE extends protection beyond the literal claims and struggles in particular with after-arising technologies, technologies that did not exist at the time of invention. Often, the after-arising technology serves as a substitute for one of the elements or limitations in a claim, thus taking the new embodiment outside the literal claims. The difficult policy question is whether the DOE should extend patent scope to cover some or all such future embodiments.
- This reflects the basic yet undecided scope issue. If patent law aims only to align incentives to invest in patentable subject matter with incentives that exist in the market but for misappropriation risks, then patent scope should not generally

doctrine. They can read a term abstractly, so that a "fastener" becomes anything that attaches two other things together, or they can read the same term more concretely, defining a fastener to be a particular type of connector such as a nail or a U-bolt. Or they may choose a level in between.

Burk & Lemley 31 (2005). Burk and Lemley suggest that "there are no hard and fast standards in the law by which to make the 'right' decision as to . . . the level of abstraction" and that "the indeterminacy is so acute that courts generally don't acknowledge that they are even engaging in [the] inquiry." *Id.* They suggest, much like Hand and many others, that the indeterminacy may be inherent in the "process of mapping words to things." *Id.* at 31, 49. Or it may be inherent in the process of legal line drawing more generally. Tribe & Dorf 1065–71 (1990). In patent law, courts obfuscate the fact that they are making such choices through the overt formalism of claims construction. Chiang calls for a more honest and deliberate approach. Chiang (2011).

[152] Burk & Lemley (2009).

> be extended to include (all) after-arising technologies. The risk of such technologies is present in markets generally; it is not special in this context. But if the economic objective of patent law is to put an additional weight on the scale in favor of such investments, that is, to provide additional incentives, then extending patent scope in this manner may be an effective way to do so. I do not make a strong claim in either direction because the analysis depends on the type of after-arising technology, whether the technology is the type that can be expected to reduce the fixed costs for entry into the market for patented invention, and the rate of fixed-cost-reducing innovation.

Patent claims describe an idea at too high a level of abstraction when the family of embodiments that would be covered extends beyond patentee's actual contribution and includes embodiments that are insufficiently related to the disclosed embodiment. This may occur when claimed embodiments either already exist in the public domain (prior art) or are not yet enabled (capable of being implemented).

The Supreme Court's analysis in *Bilski* can be understood within this framework. The Court concluded that claims 1 and 4 were not patentable because they were written at too high a level of abstraction and would convey patent scope well beyond the family of embodiments Bilski actually invented. The claims described a family of embodiments that included many preexisting embodiments and many not-yet-invented embodiments; the court described this effect as preempting the field.

The seminal "abstract ideas" case, *O'Reilly v. Morse*, involved nearly identical analysis.[153] Morse owned a patent on the telegraph, and the Supreme Court declared invalid his eighth claim, which read:

> Eighth. I do not propose to limit myself to the specific machinery or parts of machinery described in the foregoing specification and claims, the essence of my invention being the use of the motive power of the electric or galvanic current, which I call electro-magnetism, however developed for marking or printing intelligible characters, signs, or letters, at any distances, being a new application of that power of which I claim to be the first inventor or discoverer.

[153] Benson does as well. In Benson, the Court held unpatentable a method for converting numerals expressed as binary-coded decimal numerals into pure binary numerals. The Court emphasized the variance of "known and unknown uses," noting that "[t]he end use may (1) vary from the operation of a train to verification of drivers' licenses to researching law books for precedents and (2) be performed through any existing machinery or future-devised machinery or without any apparatus." As in Morse and Bilski, the claim stated an idea at too high a level of abstraction because the family of embodiments/implementations that shared that idea would extend well beyond what the patentee actually invented or enabled others to use. See, e.g., Parasidis 395 (2010); Schaafsma 432 (2010); In re Bilski, 545 F. 3d 943, 954 (Fed. Cir. 2008). See also Lemley et al. (2011).

The Court explained:

> It is impossible to misunderstand the extent of this claim. He claims the exclusive right to every improvement where the motive power is the electric or galvanic current, and the result is the marking or printing intelligible characters, signs, or letters at a distance. If this claim can be maintained, it matters not by what process or machinery the result is accomplished. For aught that we now know, some future inventor, in the onward march of science, may discover a mode of writing or printing at a distance by means of the electric or galvanic current, without using any part of the process or combination set forth in the plaintiff's specification. His invention may be less complicated—less liable to get out of order—less expensive in construction, and in its operation. But yet if it is covered by this patent, the inventor could not use it, nor the public have the benefit of it, without the permission of this patentee.
>
> Nor is this all; while he shuts the door against inventions of other persons, the patentee would be able to avail himself of new discoveries in the properties and powers of electro-magnetism which scientific men might bring to light. For he says he does not confine his claim to the machinery or parts of machinery which he specifies, but claims for himself a monopoly in its use, however developed, for the purpose of printing at a distance. New discoveries in physical science may enable him to combine it with new agents and new elements, and by that means attain the object in a manner superior to the present process and altogether different from it. And if he can secure the exclusive use by his present patent, he may vary it with every new discovery and development of the science, and need place no description of the new manner, process, or machinery upon the records of the patent office. . . . In fine, he claims an exclusive right to use a manner and process which he has not described and indeed had not invented, and therefore could not describe when he obtained his patent. The court is of opinion that the claim is too broad, and not warranted by law.[154]

Many patent scholars and judges have wondered whether the Supreme Court espoused a categorical limitation on patentable subject matter or simply a limitation on claim scope. Apart from the categorical exclusion of abstract ideas from patentable subject matter, patent law recognizes that a patent must enable others to make or practice the invention, which means making an embodiment within the family of claimed embodiments. Moreover, a number of patent doctrines do plenty of work (a) delineating what is and what is not patentable and (b) effectively shaping patent scope. For example, patent law requires novelty, utility, nonobviousness, and an enabling disclosure in a written

[154] 56 U.S. 62 (1854). The Court went on to discuss other examples.

description; each of these doctrines serve (a) and (b) and other functions as well. In the line of cases from *Morse* to *Bilski*, the Supreme Court is espousing both a categorical limitation on patentable subject matter and a limitation on claim scope. The two go hand in hand, as described below.

The analytical framework I set forth ends up in a similar place as Chiang.[155] Judges are put in the difficult position of choosing the appropriate level of abstraction to determine the scope of the patent and range of embodiments subject to the patent owner's exclusive rights. Like Chiang, I believe the scope of the claims is and should be influenced by the nature of the actual invention and the characteristics that make it novel, useful, and non-obvious. This steers away from overly formalistic claim interpretation and toward a more deliberate engagement with the underlying scope question and the context.

Theoretically, one potential advantage of my approach is that it is compatible with the notion that ideas are not patentable and only the implementation or embodiments can be patented. The Supreme Court should stop referring to *abstract* ideas and using $E = mc^2$ to illustrate and justify exclusion.[156] Focusing on this extreme and easily made case obscures the underlying issue, as does an overly formalistic approach to claims construction, where judges and the public are led to believe that claim construction inquiry is merely a linguistic exercise, a search for the "true meaning" of claim language.[157] It would be more honest, transparent, and consistent with the First Amendment and copyright treatment of ideas to make clear that ideas are not patentable. Shifting the frame in this fashion leads to a better understanding of what courts and the Patent and Trademark Office are actually doing when evaluating claims, and it also leads to a better understanding of the common task in patent and copyright and reliance on levels-of-abstraction analysis in performing that task.

I recognize that the framework I offer does not provide a clear, bright-line rule or test, and that is what many patent scholars, practitioners, and judges are looking for. In *Bilski*, the Supreme Court rejected the Court of Appeals for the Federal Circuit's exclusive reliance on the "machine or transformation" test.[158] I believe what the Court struggled to say

---

[155] It also resonates well with arguments made by four prominent patent law scholars. Lemley, Risch, Sichelman, & Wagner (2011).

[156] Lab. Corp. of Am. Holdings v. Metabolite Labs., Inc., 548 U.S. 124, 126 (2006), quoting Diamond v. Diehr, 450 U.S. 175, 185 (1981). "The principle means that Einstein could not have 'patent[ed] his celebrated law that $E = mc^2$; nor could Newton have patented the law of gravity.'" *Id.*, quoting Diamond v. Chakrabarty, 447 U.S. 303, 309 (1980).

[157] Chiang (2011).

[158] The Federal Circuit had held that "a "claimed process is surely patent-eligible under §101 if: (1) it is tied to a particular machine or apparatus, or (2) it transforms a particular article into a different state or thing." In re Bilski, 545 F. 3d 943, 954–955 (Fed. Cir. 2008). In some important respects, that test follows the analysis I have suggested, because it differentiated abstract ideas from patentable method inventions based on whether or not the method was implemented on a machine or produced a physical transformation. Thus, one might conclude the Court went in a different direction than I would support. It depends on how one interprets the decision.

is that bright-line rules are inappropriate for deciding cases at the boundary between unpatentable ideas and patentable inventions. Despite its reaffirmation of the oft-expressed and easily stated bright-line rule (abstract ideas are not patentable, which should be *ideas are not patentable*), practical reality requires a standard; it requires levels-of-abstraction analysis in context, much like copyright, albeit with a different set of doctrinal tools.

Finally, in addition to the levels-of-abstraction framework, some other related tools from copyright might be useful in patent law. For example, Mark Lemley asked me to consider patents concerning ideas that have only two different embodiments, say a product and process, each of which is claimed. Such a hypothetical may arise in the case of a patent claiming both an isolated and purified chemical compound and the sole feasible method for its isolation and purification. In such a case, it might be appropriate to import the merger doctrine from copyright law, given the possible social value of the chemical compound and the inability of competitors to design around the patented method. As noted, the merger doctrine limits copyright protection of expression when the idea being expressed is capable of being expressed in an extremely limited number of ways. One could imagine a similar merger doctrine being developed in patent law. I do not take a position on whether this would be attractive policy or whether such a patent doctrine should lead to no patent protection or thin protection. Answering this question depends on how one answers prior questions about the economic objectives or ends of patent law.

* * *

Ideas illustrate the semi-commons structure of intellectual property systems. Ideas are managed as commons by virtue of their exclusion from patent and copyright and inclusion in the public domain. Patent and copyright laws grant limited private rights over outputs from particular uses of ideas. Both patent and copyright laws struggle to filter ideas—to determine whether something is patentable or copyrightable subject matter—and to manage the scope of rights granted to avoid de facto protection of ideas. The next section explores the semi-commons structure of patent and copyright laws more generally.

## D. Intellectual Property Laws as Semi-commons Arrangements

The intellectual property laws construct semi-commons arrangements, complex mixtures of interdependent private rights and commons.[159] Semi-commons exist at different scales.

---

The Court did not turn away from the "machine or transformation" test altogether. The underlying principle (that implementation differentiates unpatentable ideas from patentable inventions) remains valid, and the test remains relevant. Courts should continue to look for machine implementation or physical transformation as an important "clue." See Bilski; Schultz and Samuelson (2011); PTO Guidelines; Lemley et al. (2011).

[159] Heverly (2003); Madison, Frischmann, & Strandburg (2010); Frischmann and Lemley (2007); Frischmann (2007b); Yu 6–8 (2005); Loren (2007); Vetter (2007). On the idea of semi-commons, see Smith (2000).

At a macro level, the cultural environment constitutes mixed infrastructure that should be managed as a commons. As discussed, commons management aims to limit both government and market shaping of the cultural environment and our lives, plans, beliefs, and preferences. Commons management is a strong default position for the cultural environment because users—autonomous individuals as well as social groups and communities—get to shape the environment and choose what to say and do and how to plan their lives, experiences, and interactions with each other and the environment.

Yet, as in the context of the natural environment, a pure open-access or commons regime can lead to tragedy, in this context associated with undersupply of certain types of intellectual resources. Consequently, intellectual property systems enclose and regulate a select (albeit very broad) set of intellectual resources. Thus, an identifiable semi-commons emerges at the macro level, with commons being the default form of management and intellectual property enclosure being exceptional, albeit of broad scope and significant importance. The unenclosed and enclosed are highly interdependent; much of which exists in either space/environment/category depends substantially on complex interactions and various inputs/contributions from the other space/environment/category. Given tremendous difficulties in establishing and maintaining boundaries, and the dynamic and complex nature of cultural-intellectual resource systems, the intellectual property laws also mediate the relationships between the enclosed and unenclosed. As demonstrated with respect to ideas, the First Amendment, copyright, and patent interact with each other and the public domain. The discussion of ideas demonstrates the semi-commons structure and associated interdependence between private and public at the macro level; at the same time, it reveals the semi-commons structure at the meso (or intermediate) level of the copyright and patent systems. Though both legal systems construct private rights and enclose a set of intellectual resources, neither constitutes pure private rights or enclosure. Rather, both copyright and patent laws themselves are semi-commons arrangements that mix both private rights and commons. Both legal systems are designed to sustain incentives and spillovers.

Copyright law creates a semi-commons arrangement—a complex mix of private rights and commons.[160] The rights granted by copyright law—specifically, the §106 rights to reproduce, display, perform, distribute, and make derivative works—provide incentives to create and disseminate works by facilitating transactions and lowering the costs of excluding competitors from using the expression. The supply-side incentives affected by copyright extend beyond the initial investment in creation to investments in content development and dissemination. What must be encouraged is not only works' creation but also their publication, dissemination, and productive use. Like traditional property rights, copyright facilitates transactions over certain uses of creative expression, and thereby enables rights holders to appropriate some of the surplus generated by their investments in

[160] The next four paragraphs derive from Frischmann & Lemley (2007), with some modification.

creation, development, and dissemination. In this fashion, the private rights component of copyright law improves investment incentives through the operation of the market mechanism; in a sense, it uses the market to achieve a broader set of economic and social ends.

The commons component of copyright law promotes spillovers; or, to put it another way, the commons component of copyright law avoids market or government allocation of resources for certain ranges of uses and for certain elements of a copyrighted work.[161] Through a variety of leaks and limitations on the private rights granted, copyright law sustains common access to and use of resources needed to participate in a wide variety of intellectually productive activities. Many of these activities generate socially valuable spillovers: benefits realized by consumers, users, and third parties that are external to a creator's decision to produce the work and to any transactions involving the work. For example, due to its limited duration, copyright has generated temporal externalities. A work that enters the public domain is free for public use, and any value derived from this use is external to both the creator's decision to produce the work and any transactions involving the work. Similarly, due to copyright's limited scope, copyright generates externalities that accrue to other creators, even competitors, as these entities can freely use various unprotected elements of a work, such as an idea, theme, or functional feature. Copyright's limited scope may also generate externalities in complementary technology markets: for example, companies can design and build products such as DVD players and iPods that facilitate the enjoyment of copyrighted works. Finally, copyright produces externalities when consumers productively use or reuse works. Creating and consuming creative expression of different types develops human capital, educates, and socializes in a manner that benefits not only creators and consumers but also nonparticipants.

Patent law, like copyright, is a semi-commons that promotes both ownership of rights and spillovers, but the particular ways in which patent law and copyright law permit "leakage" differ significantly. Patent law protections have a much shorter duration than copyright, permitting inventions to enter the public domain more quickly. Patent law also excludes some inventions from protection because requirements for obtaining protection are stricter. Once inventors do obtain protection, however, the right they obtain is much stronger and less leaky than that afforded by copyright law.

Patent law promotes spillovers in several ways. Patents generate externalities by facilitating learning and disclosure.[162] Indeed, patent law, unlike copyright law, requires the patent owner to teach the public how to make and use the invention, and this is often identified as a central function of the patent system, though in practice it is considerably

[161] Frischmann 659 (2007b).

[162] Patents may be justified as a publicly preferable means of providing exclusion when compared with alternatives such as trade secrecy. Patents lead to public disclosure of inventions and, at least some, surrounding information and ideas that would otherwise remain secret; such a justification for patents presumes a semi-commons governance regime and not pure exclusion because to be publicly meaningful, disclosure must actually enable productive uses, activities, and opportunities that would not otherwise be publicly available.

less important than the system's incentive effects. Patents lead to temporal externalities—spillovers the occur when the patent expires. Temporal spillovers are quite significant. For example, the overwhelming majority of the social benefit associated with the telephone (and, for that matter, the paper clip) occurred after the basic patents on those technologies expired.

These legal systems sustain commons by excluding resources and designating them unprotectable, but also by sustaining public access to privately owned resources for certain types of uses.[163] In a sense, the legal systems also construct semi-commons at the micro level of the protected expression or invention.[164] At this micro level, copyright law appears to be more sensitive to and accommodating of social demand for commons management of infrastructural expression than patent law is with respect to infrastructural invention. "Copyright encourages and sustains participation in intellectually productive activities that both generate and use expressive works to communicate, entertain, teach, and engage us in many different ways. Many of these activities—e.g., education, community development, democratic discourse, political participation—generate socially valuable . . . spillovers."[165] For example, fair use is a particularly important copyright law doctrine that aims to preserve public capabilities to use copyright protected expression in various ways.[166] As Lemley and I explain:

> Many paradigmatic uses deemed fair involve use of a work to engage in activities that yield diffuse, small-scale spillovers to a community. Using a work for educational purposes, for example, not only benefits the users themselves, but also, in a small way, benefits others in the users' community with whom users have interdependent relations—reading and learning builds socially valuable human capital. Critiquing a work similarly benefits not only the user but also, in a small way, others in the users' community—not only because those others may read the critique itself, but also because engaging in critical commentary is a form of creative and cultural activity that builds socially valuable human capital. We recognize that observing and measuring these spillover benefits is probably an impossible task. That is our point, in fact. As a society, on the whole, we recognize the value of active, widespread participation in these types of activities, and we know that creative expression is essential to participation. Thus, we encourage common access to and use of expression for these types of activities.[167]

---

[163] Litman (1990); Landes & Posner (2003); Gordon (1982).

[164] We might connect the idea that the infrastructure theory has a fractal nature in the sense that infrastructural characteristics manifest at different scales with the semi-commons structures (management institutions) that manifest at different scales.

[165] Frischmann 672 (2007b).

[166] Frischmann & Lemley 286–290 (2007).

[167] *Id.* at 289.

Other features of copyright law, such as constraints on the exclusive scope of rights (for example, private display and performance is permissible) or judicial willingness to provide "thinner" protection for certain types of works, also provide breathing space. The levels-of-abstraction framework described above permits judges to adjust the scope of copyright based on the context and nature of the work, and judges routinely filter infrastructural elements of a work in addition to ideas (for example, stock literary elements).[168] Patent law is not sensitive to social demand for commons management of infrastructural invention.[169] The primary commons components of patent law are its mechanisms for exclusion and conferral to the public domain (for example, disclosure, strict qualification criteria, and duration). There is no fair use or functionally equivalent doctrine. Courts implicitly engage in levels-of-abstraction analysis when construing claims and determining infringement, and that provides an opportunity to adjust patent scope, but the analysis is not sensitive to social demand or the infrastructural characteristics of the invention.[170] I do not take a position on whether or not patent law should be more sensitive to such concerns. Other scholars have advanced arguments for a patent fair use doctrine, a more robust experimental use defense, and adjusting remedies when patents on infrastructural inventions are infringed.[171]

## Appendix: Basic Research

This appendix provides an abbreviated discussion of basic research as an example of intellectual infrastructure. Before proceeding, it is important to acknowledge that basic research is itself a broad, malleable, and contested concept. There are competing definitions of basic research. In *Science: The Endless Frontier*, Vannevar Bush supplied what is now the classic definition of basic research: research "performed without thought of practical ends."[172] This and various other definitions focus on what motivates or guides researchers, what researchers expect to accomplish or hope to achieve. Some scholars

[168] There are various other examples of intellectual infrastructure excluded from intellectual property protection. See *id.* at 286 (historical facts); Lee (2008) (discussing various examples).

[169] Lee (2008). Lee provides an excellent account of how trademark, copyright, and patent law accommodate social demand for access to infrastructural works.

[170] One exception may be pioneering inventions, which may obtain broader scope because of their significance. Sun Studs, Inc. v. ATA Equip. Leasing, Inc., 872 F. 2d 978, 987 (Fed. Cir. 1989); Thomas 58–59 (1995); Lemley 1003 (1997).

[171] O'Rourke (2000) (fair use); Strandburg (2011) (same); Mueller 9–10 (2001) (experimental use); Caruso (2003) (same); Lee (2008) (exploring the concept of intellectual infrastructure, and proposing that "courts should consider the infrastructural use of a patented invention when determining infringement remedies and, in certain circumstances, allow such use to continue by a downstream user contingent upon providing compensation to the patentee.").

[172] Bush (1954).

refer to basic research as "curiosity-driven."[173] Others implicitly adopt the linear model discussed above and focus on the nonimmediacy of *commercial* applications.[174]

I would put aside what motivates the researchers (curiosity, financial returns, prestige, etc.) and whether the research is commercial or not. Simply put, many different motivations may be in play, and commerciality is not relevant to the definition of basic research (i.e., there are plenty of examples of applied noncommercial research and of basic commercial research).

Given the dual nature of research as a mixed process/product (activity/resource), I approach the definitional question by examining research from both an *ex post* and *ex ante* perspective: From an *ex post* perspective, a research output is characterized by its use, and from an *ex ante* perspective, a research project or investment is characterized in terms of its potential outcomes and uses. Thus, for each potential research activity or investment, there is a probability distribution describing the likelihood that the research will serve as an input for a range of uses. *Ex ante*, public or private investors can estimate this distribution given publicly available and privately held information (hereinafter termed the "use estimate"). The distinction between basic and applied research can be understood by looking to the *variance* of the use estimate. A larger (smaller) variance in the distribution corresponds to basic (applied) research, representing a wider (narrower) range of potential uses and hence greater (less) uncertainty as to a specific use. Therefore, on this view, the distinction between basic and applied research is not dependent on the uses themselves, that is, whether the research is commercial or not. Instead, the distinction rests on the *range of potential uses* and the corresponding *uncertainty with regard to specific uses*.[175]

Over time, moving from basic to applied research might proceed in steps, analogous to Bayesian learning, such that successful research steps affect the use estimates of subsequent research. Often, taking consecutive steps entails producing research and using it as an input to produce dependent or "second-generation" research. The linear model of the innovative process assumes a gradual narrowing of the use estimate as though deviation from the mean does not occur and learning from spillovers is a fiction. In reality, though, progress is nonlinear, meaning that distributions rise, flatten, and shift location dynamically from step to step.

Basic research seems to encompass a wide range of activities and practices among different communities (government, industry, academia, and various other actors and institutions) as well as the research results, the inputs, outputs, and so on. We might

[173] Strandburg (2011).

[174] Landes & Posner 305–06 (2003) ("Basic research is distinguished from applied research mainly by lacking *immediate* commercial applications.").

[175] Frischmann 365–66 (2000) (arguing that the difference between basic and applied research is the variance of anticipated applications or uses).

conceptualize basic research as a complex system within the cultural environment. For the sake of brevity, I will not fully explore the contours of the system or the interactions among the various resources and participant communities. Instead, having acknowledged these definitional complications, I proceed to apply the infrastructure theory to basic research.

What makes basic research valuable to society? Again, like a road system, communications networks, and oceans, basic research is socially valuable primarily because of what it facilitates downstream—how it can be used to produce further knowledge and research. It satisfies all three criteria in the general definition of infrastructure and should be classified as public infrastructure: It is nonrival; it creates benefits or value primarily because of the downstream uses, which generally involve the production of additional public goods (e.g., more research, information, knowledge, and learning); and, by my definition, there is wide variation in downstream uses. Again, in my view, what distinguishes basic research from applied research is the variance in expected (or, in some cases, desired or predicted) uses of the research.[176]

It is difficult to estimate the social value of basic research, primarily because of the wide variety of downstream uses that generate public goods and uncertainty with respect to future directions that the cumulative productive processes may go. Basic research, like many infrastructural resources, builds public capabilities to be productive in various ways that are difficult to fully trace. Nonetheless, as with many traditional infrastructures, it is well recognized that basic research contributes significantly to economic growth and social welfare.[177]

As with other infrastructure, recognizing that basic research behaves economically as infrastructure suggests that the social costs of restricting access to the resource can be significant and yet evade observation or consideration within conventional economic transactions. Many others have noted that granting exclusive property rights (e.g., patents) over basic research[178] stifles downstream research, which can impose substantial social costs.[179] This does not mean that no progress will be made. Some avenues of follow-on research may proceed, for example, by initial researchers or others to whom licenses are granted. The point is that basic research may "be encumbered with excessive licensing

[176] There are obviously differences between desired, expected, and predicted, but since basic research is rather amorphous and continuous and has a dual product/process nature, I am reluctant to draw any sharp lines.

[177] See, e.g., Landes & Posner 305–08 (2003); Rai (1999); Reichman & Uhlir (2003).

[178] While a significant amount of basic research is not patentable, it appears that "more and more fruits of basic research [can] be patented," Landes & Posner 308 (2003). In some areas, at least, both the existence and the prospect of patents have had a significant effect on the research process. See *id*. at 305–08.

[179] See *id*.; Scotchmer 32 (1991); Merges & Nelson 869–80 (1990). As Robert Merges and Richard Nelson explain, some private firms recognize the value of open access to basic research and have undertaken efforts to place research results in the public domain. *Id.*

fees and transaction costs,"[180] and the paths taken may be unduly constrained from a social perspective.

Moreover, granting property rights over basic research links resource management with commercialization and introduces the market mechanism's inherent bias for outputs that generate observable or reasonably foreseeable and appropriable returns.[181] Thus, in making decisions regarding access, owners would face the same set of problems that the hypothetical owner of a lake in chapter 11 would face—for example, excessive transaction costs and uncertainty regarding the prospect of appropriable returns.[182] While downstream uses of basic research are not rivalrous in the technical sense (i.e., there is no risk of congestion), users may compete with each other to develop and commercialize the research and may demand exclusive licenses.[183] In granting such licenses, owners may favor uses reasonably expected to generate appropriable returns at the expense of uses more likely to generate positive externalities.[184] This may retard progress in a manner that has substantial social opportunity costs in the sense that socially valuable research paths lie fallow and unexplored.[185]

Consider the case of basic research that has uncertain or low commercial value, which, according to Arti Rai, deserves particular attention:

> [I]n the context of research that is demonstrably of low commercial value, there is evidence that upstream proprietary rights have impeded downstream research.

---

[180] Merges (2004).

[181] Not only does this bias affect management of existing research results; it also has dynamic effects on the research process because the prospect of obtaining a patent may skew researchers' incentives and basic scientific norms. See Frischmann (2009a); Lee (2009); Rai 109–13 (1999); Frischmann (2005c); see also Scotchmer 127–31 (2004). Scotchmer explains: "[I]t is not easy to compensate the developers of basic technologies. Commercial value generally resides in products that are developed later. If the founders earn some profit, it is only because they can demand licensing fees from later developers. But this requires that later products infringe their patents. Basic scientific knowledge . . . is generally not patentable, in recognition of the fact that the benefits would be hard to appropriate."
*Id.* at 129. One reason that basic research should be supported by public sponsors rather than private investors "is that the benefits of basic research are hard to appropriate by private parties." *Id.* at 131–32. To the extent that the public goods applications are sufficiently commercializable (applied and commercial), there is an argument that markets should work quite well in manifesting demand for the infrastructure and that the major impediments to maximizing social welfare originate on the supply side. See *id.* at 127–59.

[182] See chapter 11.

[183] Notably, some of the downstream markets involve further investment in public goods as the innovative process continues.

[184] Cf. Rai (2005) ("[I]n university contexts, where the immediately foreseeable payoffs—commercial or academic—from research is often not high, researchers are unlikely to be willing or able to incur high transaction costs in order to gain access to upstream research.").

[185] In an earlier article, I argued that this constitutes a special type of market failure, which I called "innovative process market failure," because the failure to pursue potential avenues of research involves hidden costs associated with the cumulative, nonlinear nature of the innovative process. Frischmann 374 (2000).

> Consider the case of research into a malaria vaccine. The disease burden associated with malaria is very significant, on the order of over one million deaths a year. The social value of a malaria vaccine would therefore be quite high. Nonetheless, because the primary market for such a vaccine would be in the developing world, such research is of low commercial value. . . .
>
> . . .
>
> . . . In the area of agricultural biotechnology, there is perhaps even more compelling evidence that research projects of low commercial value have been significantly delayed, or have not gone forward at all, because of upstream patent rights. Specifically, restricted access to patented technologies has been identified as a significant barrier to development of subsistence crops relevant to the developing world.[186]

More generally, the social costs associated with the market mechanism's inherent bias for outputs that generate observable and appropriable returns may be significant. These costs evade observation because basic research is often an input into and output from cumulative processes involving multiple inputs, multiple outputs, multiple actors, and multiple research avenues heading in different directions. These cumulative processes involve nonlinear progression, feedback loops, (cascading) spillovers, and numerous other complications that frustrate modelers and defy simplification.[187] All of these characteristics contribute to information and transaction cost problems that make relying on property-based, market-driven management of basic research results almost outrageous, much like the seemingly ridiculous hypothetical (in chapter 11) of granting ownership of Lake Michigan to an individual property owner.

Edmund Kitch's "prospect theory" of patents simply does not work well for basic research.[188] His theory is premised on two notions: (1) that the property owner will minimize social waste associated with duplicative efforts; and (2) that the property owner

---

[186] Rai (2005). Rai provides a number of specific examples where upstream patents have impeded downstream progress of research with low commercial value. See *id.* Rai also considers whether collective action may alleviate the problem. See *id*.

[187] Consideration of these characteristics is beyond the scope of this chapter. There is, however, substantial literature in this area. See, e.g., Scotchmer (2004); Scotchmer (1991).

[188] See Kitch 265, 276–78 (1977) (likening a patent to a mining claim, where society benefits from efficient exploitation of a patent and its prospects). Kitch argued that IP rights facilitate efficient coordination of research such that duplicative waste is minimized. There are a number of scenarios where this argument holds true, some of which depend on the bargaining positions of primary and secondary researchers, and some of which depend on the innovation types involved. See, e.g., Green & Scotchmer 20, 31 (1995) (discussing bargaining positions of primary and secondary innovators in IP licensing context). Perhaps the most straightforward example of the bargaining position scenario is where valuable information is asymmetrically held by the primary researcher that can be selectively doled out to willing licensees. The latter scenario may occur, for example, in the derivative market for applied incremental research for which duplication is likely, i.e., the range of improvements is narrow.

will best commercialize and license an invention.[189] Neither premise, however, holds up with respect to basic research. Wasteful duplication seems much less likely to be a problem in the context of basic research because of the multitude of directions and research paths that grow out of basic research. One must adopt a rather strong version of the linear model of innovation to support the premise that competitive research efforts within and building from basic research will lead to wasteful duplication.[190] This view assumes that research efforts follow the same course of research, lead to identical results, and are thus economically wasteful. But such an assumption is valid only for relatively applied research projects; two independent research projects having the same extremely peaked use estimate will likely result in duplication because achieving the expected mean use is highly likely. But the likelihood of wasteful duplication diminishes as the variance increases.

- Assume that firms A and B each begin with a shared pool of common knowledge and identical sets of resources (capital, labor, expertise, know-how, etc.). Furthermore, each begins an independent research project X at time $t_1$ with identical use estimates $P_1(X)$ but without any sharing of information after $t_1$. The outcome of each firm's project at $t_2$ yields $X_A$ and $X_B$. The likelihood that the outcomes are the same depends on the variance of the estimate. When the outcomes are secret and not identical, each firms' new use estimate for continued research is different, creating both different opportunities for progress in the broad sense and different incentives for continued investment. The communication of outcomes from period 1 leads to a separate use estimate.[191]

[189] Kitch 276–78 (1977). See Frischmann 372–73, 374–76 (2000); Merges 359, 381 (1992); Scotchmer 155 (2004). Scotchmer concludes:

> Thus the licensing platform created by a pioneer patent can undermine competition . . . in the "innovation market" . . . and competition among users of the patented knowledge. It might be better not to give such patents. One alternative is public funding, and another is to let a later innovator who needs the pioneer innovation redevelop it. This leads to cost redundancy, but unless the tool is very expensive, such redundancy may be a lesser evil than retarding the development of later products through restrictive joint ventures or raising their price by facilitating collusion.

*Id*.

[190] In a discussion of patent races, Scotchmer associates inefficient duplication with the assumption that "R&D costs have a large fixed component [where] if two firms invest, the cost is needlessly duplicated." Scotchmer 273, 275 (1998). She distinguishes the fixed R&D cost process, which leads to duplicative results, from a variable R&D cost process, which leads to duplicative but accelerated results. The latter may be regarded as efficient (i.e., nonwasteful) duplication.

[191] Frischmann (2000).

Various researchers participating in basic research may hasten progress or lead to advances in different directions.[192] If the linear model is dropped, there is little reason to believe that coordination of basic research is akin to mining for ore.

Managing basic research as a commons may have considerable appeal as a means to leverage nonrivalry, maintain the social option value of basic research, and encourage widespread productive use. But how do we overcome the production problem? How do we overcome supply-side problems for basic research?

Basic research is both publicly and privately provided, and there is a continuum of public, private, and hybrid institutions that address supply-side concerns, including grants, procurement, subsidies, regulation, property rights, intellectual property rights, contracts, tax incentives, technology, and social norms. Notably, intellectual property is much less of a factor than public funding. According to William Landes and Richard Posner:

> An enormous amount of basic research is produced every year in the United States and other advanced countries without benefit of patentability. . . . In 1999 half of all basic research in the United States was funded by the federal government, and of the balance 29 percent was financed by universities and other nonprofit research establishments out of their own funds.[193]

By 2010, the portion of basic research funded by the federal government had grown to 60 percent, with academic institutions continuing to act as the second largest source of funding.[194] Universities, many of them publicly-sustained, conduct 55 percent of basic research, with business and industry accounting for less that 20 percent.[195]

Public financing reduces the need to rely on private investment and eliminates supply-side concerns over protecting incentives to invest in the research. In theory, it would appear that the optimal management decision would be to release publicly funded basic research results into the public domain to encourage free, widespread, and potentially competitive use downstream. In a sense, public investment in a basic research commons is precisely the sort of indirect intervention that does not aim for optimality (or optimal

---

[192] As Merges and Nelson argue, "rivalry facilitates technical advance and unified control damps it." See Merges & Nelson (1990). As Richard Nelson put it, "From a social point of view, effective pursuit of technological advance seems to call for the exploration of a wide variety of alternatives and the selective screening of these [alternatives once] their characteristics have been better revealed a process that seems wasteful with hindsight." Nelson 455 (1982).

[193] Landes & Posner 306 (2003).

[194] The Science Coalition, Sparking Economic Growth 3 (2010), available at http://www.sciencecoalition.org/successstories/resources/pdf/Sparking%20Economic%20Growth%20Full%20Report%20FINAL%204–5-10.pdf (arguing for increased public investment in basic research, along with increasing university-based research, and pointing out the positive economic benefits and job creation that such investment would engender).

[195] *Id.*

investment in public goods production)[196] but instead aims to support a wide range of "follow-on" activities, by enhancing basic public capabilities,[197] improving the knowledge base on which the public can build, and maintaining flexibility in the opportunities available.

In reality, as one would expect, there are serious obstacles to effectively implementing this solution. To begin with, it depends on government to raise sufficient funds and allocate them efficiently. Of course, government financing of infrastructure always raises these issues, but basic research funding raises some distinct issues. First, in comparison with most traditional infrastructure, basic research is much less tangible and visible to citizens, who may be asked to pay higher taxes to fund it. Second, the demand-measurement problem highlighted by Samuelson may be especially relevant if we substitute tax-paying communities for consumers; communities have strong incentives to conceal their preferences in the hope of free riding on the investments of other communities. Thus, it is not surprising that very little basic research funding comes from local or state governments.[198] Even at the national level, we might expect systematic underinvestment in basic research because of free-riding concerns. The United States is the leading supporter of R&D activities in the world, accounting for 33 percent of worldwide R&D expenditures.[199] Public support for government funding of basic research remains strong, with 84 percent of Americans expressing support for such funding in 2008 (at the height of a long and painful recession).[200]

Another major obstacle is effective management of basic research. In fact, based in part on the perception that the federal government had a very poor record of managing federally funded research results,[201] Congress enacted a series of legislative reforms, such as the Bayh-Dole Act,[202] that aimed to facilitate the transfer of publicly funded technology to the private sector.[203] Most notably, the Bayh-Dole Act permitted and encouraged

[196] The demand-measurement problems truly would make such an effort impossible.

[197] Education is the other obvious example.

[198] General R&D funding from nonfederal sources "is small in comparison to federal and business sources"; in 2008 the combined funding from state and local governments, along with academic institutional funds, and nonprofits was a mere 7 percent. National Science Foundation, Science and Engineering Indicators 2010 at 4–14 (2010), available at http://www.nsf.gov/statistics/seind10/pdf/c04.pdf.

[199] National Science Foundation, Science and Engineering Indicators 2010 at 4–33 (2010), available at http://www.nsf.gov/statistics/seind10/pdf/c04.pdf. Japan is the second-largest performer at 13 percent, with China in third at 9 percent of global R&D expenditure. *Id.*

[200] *Id.* at 7–29.

[201] See Eisenberg (1996) (explaining and critiquing this perception).

[202] See Bayh-Dole University and Small Business Patent Procedures Act, Pub. L. No. 96-517, 94 Stat. 3019 (codified as amended at 35 U.S.C. §§ 200–211) (2000); see also Stevenson-Wydler Technology Innovation Act of 1980, Pub. L. No. 96-480, 94 Stat. 2311 (codified as amended at 15 U.S.C. §§ 3701–3714 (2000)).

[203] On these legislative reforms, see Eisenberg 1704–09 (1996); Eisenberg (1994); Frischmann 406 (2000); Rai 92–94, 109–15 (1999).

federally funded researchers to obtain patent rights over their inventions. The premise was that patents would facilitate postpatent research, development, and commercialization. That is, in the absence of patents, government-funded research results would languish underutilized because (1) the researchers and their host institutions lacked the incentives and/or capacity to further develop and commercialize the research or to transfer the research results to industry, and (2) even if transfer was feasible, industry lacked sufficient incentives to invest in development and commercialization without the exclusivity made available by patents in the form of exclusive licenses. Granting researchers patent rights, it followed, would enable them to better manage their inventions, and would encourage cooperation between university researchers and industry.[204] Relying on intellectual property to stimulate technology transfer reflected a fundamental shift within the university research community.

The shift has had a profound effect on basic research efforts in certain fields. For example, as noted by Walter Powell, there has been a "sea change in the focus of basic research" in life sciences because of commercialization by universities of basic scientific research results.[205] Moreover, the effort to bring academia and industry closer together may have had significant effects on universities and their science and technology research systems.[206] It is very difficult to gauge this type of institutional change, but there are some indications that the Bayh-Dole Act has had impacts on the management of university science and technology research systems. The *Economist* magazine, which in 2002 heralded the Bayh-Dole Act as "[p]ossibly the most inspired piece of legislation to be enacted in America over the past half-century,"[207] more recently concluded:

> Many scientists, economists and lawyers believe the act distorts the mission of universities, diverting them from the pursuit of basic knowledge, which is freely

---

[204] See Eisenberg 1664–66 (1996).

[205] Powell (2001); see also Eisenberg 223 (2001) (suggesting that delays and high transaction costs stifle transfers of biotechnology research tools).

[206] Many have documented the significant increase in commercial activities of universities, including patenting and licensing, for example. There are many different explanations, however. See generally Slaughter & Leslie (1997) (studying multiple policy instruments and their commercialization impact). In fact, a number of scholars "have argued that much of the increase in commercially oriented university activities, such as patenting and licensing, that has occurred since 1980 was driven by contemporaneous shifts in intellectual property laws and regimes for funding academic research." Shane (2004) (citing Henderson, Jaffe, & Trajtenberg 119 (1998); Mowery & Ziedonis 399 (2002); Mowery et al. 99 (2001); see also Mowery 16 (2005). Mowery shows that the trend of increased patenting behavior by universities occurred prior to 1980 and the passage of Bayh-Dole. He suggests that, while the relationship between universities and industry may have evolved (been transformed) in the past few decades, transformation should not be attributed to the Bayh-Dole Act itself. *Id.* On university science and technology research systems as infrastructure, see Frischmann 2143 (2009) (applying infrastructure theory to university science and technology research systems and explaining how patents enable a demand pull on the allocation of university research system resources).

[207] *Opinion, Innovation's Golden Goose*, at 3.

> disseminated, to a focused search for results that have practical and industrial purposes. Whether that is a bad thing is a matter of debate. What is not in dispute is that it makes American academic institutions behave more like businesses than neutral arbiters of truth.[208]

Despite various expressions of concern about these types of impacts, the empirical evidence is rather light, in part because institutional change may be slow, subtle, and difficult to measure empirically.

There is a rich literature on the impact of the Bayh-Dole Act and related policies, and my objective here is not to engage in that debate. Instead, let me suggest that for basic research, coupling government funding with a clear dedication to the public domain remains a potentially attractive method for sustaining a commons that relies on neither the government nor the market mechanism to manage access. Yet it may not be enough to "release basic research results into the public domain." It is not as if the public domain is the atmosphere that subsequent users can simply inhale and use. Dissemination and effective transfer of basic research to a community of users often requires additional steps beyond mere "release" to the public domain. What those steps may be depends on the context; in some cases, publication of research results in peer-reviewed journals is critical; in other cases, collaboration among researchers in academia and industry may be critical; in some cases, an open-access repository might suffice.

[208] *Bayhing for Blood* 109 (2005) ("For example, a study published in 2003 by Jerry and Marie Thursby, of Emory University and the Georgia Institute of Technology respectively, showed that more than a quarter of the licenses issued by universities and research institutes include clauses allowing the business partner in the arrangement to delete information from research papers. Almost half allow them to insist on publication being delayed."); Reichman & Uhlir 341 (2003) ("Under Bayh-Dole, universities have moved away from policies that favor pure research, both for its own sake and as a tool for advancing higher education. As the costs of education skyrocket, and government funding fails to keep up in many areas, universities have aggressively sought to exploit commercial applications of research results, with an eye toward maximizing returns on investment."). See generally Kesan (2009) (collecting sources).