



Infrastructure and general purpose technologies: a technology flow framework

Christiaan Hogendorn¹ · Brett Frischmann²

Published online: 18 February 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Studies of economic growth often refer to “general purpose technology” (GPT), “infrastructure,” and “openness” as keys to improving productivity. Some GPTs, like railroads and the Internet, fit common notions of infrastructure and spawn debates about openness, such as network neutrality. Other GPTs, like the steam engine and the computer, seem to be in a different group that is more modular and open by nature. Big data, artificial intelligence, and various emerging smart technological assemblages have been described both as GPTs and infrastructure. We present a technology flow framework that clarifies when a GPT is implemented through infrastructure, provides a basis for policy analysis, and defines empirical research questions. On the demand side, all GPTs—whether implemented through infrastructure or not—enable a wide variety of productive uses and generate substantial spillovers to the rest of the economy. On the supply side, infrastructure is different from many other implementations of GPTs; infrastructure is partially nonrival, which may complicate appropriation problems and raise congestion issues. It also exhibits tethering, meaning that different users must be physically or virtually connected for the infrastructure to function, and this makes control of its uses more feasible and more salient to policy.

Keywords Infrastructure · General purpose technology (GPT) · Institutions · Economic growth · Spillovers · Open technology

JEL Classification H54 · K23 · N70 · O30

✉ Christiaan Hogendorn
chogendorn@wesleyan.edu

¹ Economics Department, Wesleyan University, Middletown, CT, USA

² Charles Widger School of Law, Villanova University, Villanova, PA, USA

1 Introduction

While there is widespread agreement that “infrastructure,” “general purpose technology” (GPT), and “openness” of technology are important to economic welfare, the connections between these concepts are often murky. We propose a conceptual framework that makes these terms clear and provides a consistent method to construct economic production functions and to motivate empirical research. The term ICT, for example, is widely used to group together computer hardware, software, the Internet, and telecommunications, but it also obscures the complex interrelationships between these technologies. Our technology flow framework gives a rigorous method to choose which of these should be labeled as infrastructure and/or as a GPT.

The most important policy contribution of our framework is to clarify the meaning and purpose of “openness” for a technology or infrastructure facility. Confused definitions muddle debate and prevent rigorous analysis in recent controversies surrounding network neutrality on the Internet, compatibility of different music and video players, and interconnection of telephone, electricity, and transportation networks. Participants in these debates frequently ask whether efficient use of infrastructure or technology can proceed from market forces alone, requires government intervention, or whether a third way involving commons management is possible. While our framework does not dictate a specific policy outcome, it does focus discussion on the precise elements of the technological system that require policy attention.

Most of the resources we think of as traditional “infrastructure,” including electricity, railroads, and the Internet, are also included in lists of what economists call general purpose technologies (GPTs). But economists typically count other technologies among GPTs—such as the steam engine and the computer—that are not usually associated with “infrastructure.” These labels matter. For example, the FCC used these ideas to justify its *2010 Open Internet Order* on network neutrality: “Like electricity and the computer, the Internet is a ‘general purpose technology’ that enables new methods of production that have a major impact on the entire economy. The Internet’s founders intentionally built a network that is open, in the sense that it has no gatekeepers limiting innovation and communication through the network.” (FCC 2010, p. 5) The FCC quoted this line and reconfirmed the argument in its *2015 Open Internet Order* (FCC 2015).

This paper untangles the relationships between infrastructure and GPTs by combining two approaches from the literature on commons and commons management. From Hess and Ostrom (2003), we use a setup where information goods (and we suggest other goods) are based on an *idea*, packaged as *artifacts*, and distributed through *facilities*. From Frischmann (2012), we use a “demand-side theory of infrastructure” in which the *demand-side uses* of infrastructure are equal in importance with its supply-side provision. Combining these ideas, we explain that, on one hand, all general purpose technologies, by definition, have extensive demand-side scope across the economy, and on the other hand, only some supply-side implementations of general purpose technologies have the scale and design

of facilities to qualify as infrastructure. When the facilities that implement a GPT qualify as infrastructure, there are greater challenges to openness of the technological system. Thus, policy-makers must take account of whether or not a GPT is implemented as infrastructure.

In the next section, we describe the sometimes-confusing terminology and concepts around infrastructure and GPTs. In Sect. 3, we build a technology flow framework for looking at the supply–demand relationships between technologies, and in Sect. 4, we give two important examples. In Sect. 5, we use the framework to specify the differences and similarities of infrastructure and GPTs, and in Sect. 6, we use the framework to address the openness question. In Sect. 7, we present some additional implications of using the technology flow framework, and we conclude in Sect. 8.

2 Infrastructure and general purpose technology

The facilities we call “infrastructure” are key to development and growth today. Infrastructure enables, drives, and interacts with trends such as globalization, integration of economies, and outsourcing. The term generally conjures up the notion of a large-scale physical resource made by humans for public consumption. Standard dictionary definitions refer to the underlying framework or foundation of a system. Familiar examples include (1) transportation systems, such as highway systems, railway systems, airline systems, and ports; (2) communication systems, such as telephone networks and postal services; (3) governance systems, such as court systems; and (4) basic public services and facilities, such as schools, sewers, and water systems. This list could be expanded considerably.

Yet economics does not have a generally accepted definition of infrastructure, nor is there a (sub)field focused on infrastructure economics per se. Following the suggestion of the National Research Council (1987), Frischmann (2012) argues a more capacious understanding of infrastructure is needed for functional economic analysis. According to Frischmann (2012), infrastructure satisfy the following criteria:

- (1) The resource may be consumed nonrivalrously for an appreciable range of demand.
- (2) Social demand for the resource is driven primarily by downstream productive activities that require the resource as an input.
- (3) The resource may be used as an input into a wide range of goods and services, which may include private goods, public goods, and social goods.

The first criterion captures the consumption attribute of public and impure public goods. In short, this characteristic describes the ‘sharable’ nature of infrastructural resources. Multiple users can access and use infrastructural resources at the same time. Infrastructural resources vary in their capacity to accommodate multiple users, and this variance in capacity differentiates nonrival (infinite capacity) resources from partially nonrival (finite but renewable capacity) resources (Frischmann 2012). The second and third criteria focus on the manner in which infrastructure creates

social value. The second criterion emphasizes that infrastructures are capital goods that create social value when utilized productively downstream and that such use is the primary source of social benefits. Societal demand for infrastructure is *derived demand*. The third criterion emphasizes both the variance of downstream outputs (the genericness of the input) and the nature of those outputs, particularly public goods and social goods.

Infrastructural resources are intermediate capital resources that often serve as critical foundations for productive behavior within economic and social systems. Infrastructural resources effectively structure in-system behavior at the micro level by providing and shaping the available opportunities of many actors. In some cases, infrastructural resources make possible what would otherwise be impossible, and in other cases, infrastructural resources reduce the costs and/or increase the scope of participation for actions that are otherwise possible. In addition, infrastructural resources structure in-system behavior in a manner that leads to spillovers. That is, infrastructural resources facilitate productive behaviors by users that affect third parties, including other users and even non-users of the infrastructure. The third-party effects often are accidental, incidental, and not especially relevant to the infrastructure provider or user. Thus, the *social returns* on infrastructure investment and use may exceed the *private returns* because society realizes benefits above and beyond those realized by providers and users (Steinmueller 1996; Frischmann and Lemley 2007).¹

Of course, various fields of economics recognize that infrastructure is *special*. Macroeconomics recognizes that infrastructure is important for economic development and a key ingredient for economic growth (Aschauer 2000; Ghosh and Meagher 2004).² Microeconomics recognizes that infrastructural resources often generate substantial social gains and that markets for infrastructure often fail and call for government intervention in one form or another (David and Foray 1996; Justman and Teubal 1995). Increasingly, there is recognition that market failures arise on both the supply and demand sides and demand-side failures also create obstacles for government provisioning (Frischmann 2012; Frischmann and Hogendorn 2015). Development economics, which applies both micro- and macroeconomics to the process of development, recognizes similar ideas. Yet there are some gaping holes in our understanding of how infrastructural resources generate substantial social gains and contribute to development and economic growth. In particular, the importance of *open infrastructure* and *user-generated spillovers* at the microeconomic level and

¹ While we are primarily focused on positive third-party effects, we note that negative effects are also possible, for example when transportation infrastructure facilitates illegal movements of goods or when Internet infrastructure facilitates divisive or mental-health-impairing forms of social media. Where such effects exist, they further heighten the pressure for some form of overarching commons management of the infrastructure.

² With respect to national and regional economic growth, Gramlich (1994, p. 1177) defines infrastructure as “large capital-intensive natural monopolies such as highways, other transportation facilities, water and sewer lines, and communication systems,” and Grimes (2014, p. 333) expands it to “capital-intensive investments that service multiple users.” Frischmann (2012) develops the idea of infrastructural resources that are a special type of public good that has three characteristics.

Table 1 List of transforming GPTs from Lipsey et al. (2005, p. 132)

Domestication of plants	Three-masted sailing ship	Motor vehicle
Domestication of animals	Printing	Airplane
Smelting of ore	Steam engine	Mass production, continuous process, factory
Wheel	Factory system	Computer
Writing	Railway	Lean production
Bronze	Iron steamship	Internet
Iron	Internal combustion engine	Biotechnology
Waterwheel	Electricity	Nanotechnology

their relationships to technological innovation and growth at the macroeconomic level deserve attention (Helpman 1998).

A recurring political and policy issue concerns who controls the conditions under which infrastructure can be used. Here we label this the “openness” question, and it relates to how much control the owner of infrastructure can exercise over particular types of use of the infrastructure and, in some cases, over users themselves. Currently, the most active debate concerns local broadband Internet and goes by the name “network neutrality.” But the same types of questions have occurred in railroads, telephone, and most other types of networked infrastructure (Hogedorn 2005). Frischmann and Selinger (2018) explain how the questions will continue to arise for a host of supposedly “smart” infrastructure systems.

The concept of *general purpose technology* (GPT), introduced by Bresnahan and Trajtenberg (1995), has become important to economic thought. Bresnahan and Trajtenberg argued that GPTs have three main features: pervasiveness, technological dynamism, and innovational complementarities, and they highlighted three examples, the steam engine, the electric motor, and the semiconductor. The Bresnahan and Trajtenberg paper sits on the cusp between microeconomics and macroeconomics, but the majority of follow-on research tilts to the macroeconomic side, particularly in the area of growth theory.

In a recent paper on GPTs, Bekar et al. (2018, p. 1012) present their preferred definition of a GPT: “A GPT is a single technology, or closely related group of technologies, that is widely used across most of the economy, is technologically dynamic in the sense that it evolves in efficiency and range of use in its own right, and complementary with many downstream sectors where those uses enable a cascade of further inventions and innovations.” The most far-reaching of these also have a transforming effect on society at large. In their book on GPTs, Lipsey et al. (2005, p. 132, Table 5.1) propose a list of these transforming GPTs (Table 1).

We contend that of these GPTs, the railway, electricity, and the Internet are often put in lists of “infrastructure,” and that the road, the seaport, and the airport are also usually categorized as infrastructure and serve as essential facilities for the wheel, the motor vehicle, sailing and steam ships, and the airplane. This brings up a difficulty in the definition of GPTs, since “the railway” may refer to the tracks, the locomotives and freight cars, or the whole system. The concept of GPTs and lists such as the above have been criticized by Field (2011) for a lack of parallelism.

The definition from Bekar et al. (2018) is designed to remedy any lack of clarity by declaring that each GPT is potentially a *group* of technologies and that a GPT is used widely across the economy whether for many different purposes (e.g. electricity) or for the same generic purpose (e.g. the three-masted sailing ship). Compiled in 2005, the list above included the emerging GPTs of biotechnology and nanotechnology. Today the nexus of artificial intelligence and big data seems to be an emerging GPT that refers to a group of complementary technologies (increasingly) used across the economy to drive automation and decision-making for many different purposes.

Helpman and Trajtenberg (1998) emphasize the need for parallel development of the components of a GPT, for example that motor vehicles require highway systems. They model diffusion of GPTs by endogenizing the development of the components of GPTs in many sectors. In their model there are multiple equilibria, with all, some, or none of the sectors coordinating around a new GPT. They do not, however, make any distinction as to whether the components are infrastructure or not.

The idea of parallel development of components is related to “poverty traps,” which originates with Rosenstein-Rodan (1943). He said that the high fixed costs of modern technology require a “big push” by government to move multiple sectors of the economy to advanced technologies. Murphy, Shleifer, and Vishny (1989) modernized the idea and laid down two key conditions for a poverty trap: (1) investment increases the size of other firms’ markets or increases the profitability of investment, and (2) investment has negative net present value. It is possible that a poverty trap may involve infrastructure, in which case the infrastructure is complementary to other sectors but has negative net present value when produced by a private firm due to its extensive spillovers (Frischmann and Hogendorn 2015).

Despite the importance of diffusion in the GPT literature and the presence of so many infrastructures or components of infrastructures among lists of GPTs, the precise relation of infrastructure and GPTs has not been specified. There also has not been much discussion of the openness of GPTs to new uses, rather it is usually assumed even though, as we shall discuss, there can be barriers. We believe this is primarily because most GPT research proceeds using macroeconomic growth methods, while infrastructure is treated as a microeconomic question for regulatory policy. We build a framework that unites the two, essentially, where micro meets macro.

3 The technology flow framework

We have seen that “infrastructure” is imprecisely defined and that GPT is precisely defined but subject to some confusion regarding categorization. We use two tools to relate and refine these concepts: the economics of the commons and the demand side theory of infrastructure.

Hess and Ostrom (2003) model information goods using techniques from the literature on commons and common pool resources. They discuss three important

aspects of an information good: the artifact, the facility, and the idea (p. 129). They define an *artifact* as “a discreet, observable, nameable representation of an idea,” and give many examples including books, maps, and web pages. Artifacts are the physical (or virtual) resource flow units in an information commons. A *facility* “stores artifacts and makes them available” and is analogous to a commons’ *resource system*. Examples include libraries, archives, and online repositories. *Ideas* are “the creative vision, the intangible content, innovative information, and knowledge” contained in an artifact.

The triplet of idea, facility, and artifact is a compelling way to approach *any* resource system, not just those dealing with information goods. Arthur (2009) defines technology simply as “a means to fulfill a human purpose”, and Lipsey et al. (2005, p. 58) say that “Technological knowledge, technology for short, is the set of ideas specifying all activities that create economic value.” Both these definitions suggest that a technology is an example of Hess and Ostrom’s *idea*. Any given implementation of a technology then requires various forms of capital (including both physical capital and intangible capital) that together constitute the *facility*.³ Finally, the facility generates the final output, which is the *artifact*. To take one example, the public utility power system (which is often what is meant by the term “electricity”) involves the *idea* of technologies using electromagnetism, the *facility* of the power grid, and the *artifact* of electrons moving through the grid measured in kilowatt-hours.

Our other tool comes from Frischmann’s (2012) book, *Infrastructure: The Social Value of Shared Resources*. He emphasizes the importance of a “demand-side theory of infrastructure” in addition to the more traditional focus on the supply-side. He argues that infrastructure is a partially nonrival input into a wide variety of productive activities that generate private, public, and nonmarket goods. The extent of partial nonrivalry is important—the infrastructural facility has large nonrival capacity relative to market demand, although on-peak demand can still create congestion and thus rivalry.⁴

Infrastructure is a “shared means to many ends” (Frischmann 2012, p. 4). For example, a road can be used by cars, trucks and motorcycles, which we can think of as analogous to Internet packets. It is on the demand side that these different vehicles (and different origin–destination pairs, different cargoes, etc.) create diverse sources of value.

Our model of a technology uses the supply/demand distinction from Frischmann. Our main contention is that both the supply and the demand sides can be divided, separately, into the three Hess–Ostrom elements of idea, facility, and artifact. Supplying an implementation of a technology requires the nonrival idea for the technology, it requires an enabling facility that is rival or partially nonrival, and it then produces an artifact which is a rival, private good that we denote x_1 . A demand-side

³ The facility is a specific part or group of parts of what Lipsey et al. (2005) call the *facilitating structure* of technology.

⁴ Frischmann also discusses purely nonrival “intellectual infrastructure,” but our model in this paper does not aim to incorporate those examples.

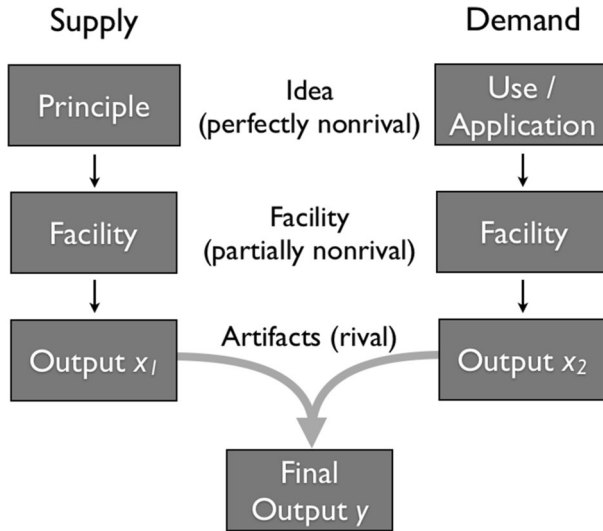


Fig. 1 The technology flow framework

use of the technology also requires a nonrival idea for a use or application, a facility to enable use, and an output which is a rival, private good denoted x_2 . In the end, the two artifacts are combined into a useful, final output y , and this combination is itself an artifact.

The technology flow framework maps directly to the traditional production function $y=f(x_1, x_2)$, where the form of the function $f()$ describes the mechanism for the combination of x_1 and x_2 . That mechanism, in turn, is usually determined by the nature of the facilities involved, as we discuss in detail below. We illustrate the technology flow framework in Fig. 1.

It is easiest to understand this framework through examples, which we proceed to shortly. But let us note three important details before moving on.

First, we can use the technology flow framework to evaluate *any* technology, no matter whether anyone would call it general purpose or call it infrastructure or not. What makes a technology “general purpose” is its many uses, and as Frischmann notes, infrastructure also has a great many uses. There are also special-purpose technologies which have fewer demand-side uses, and indeed Lipsey et al. (2005) note that there are many “almost-GPTs” that are important technologies but not used widely enough across the economy to qualify as GPTs. As long as a technology has at least one use, it can fit in the technology flow framework.

Second, every technology is in fact an interdependent combination of other technologies, as many authors, including Rosenberg (1978) and Arthur (2009), have made clear. That means that one can nest the technology flow framework over and over, and in multiples, as in Fig. 2. Every use of one technology can become a technology in itself that leads to other uses, and so on in long progressions. Even GPTs fall into this category; for example, the technologies that make up the railway are demand side uses of the wheel, the steam engine, and the internal combustion

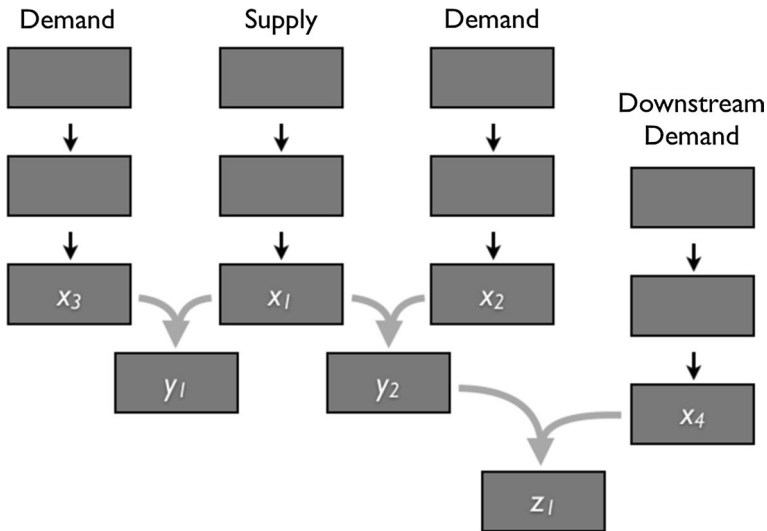


Fig. 2 Extended technology flow relationships

engine. As with any model, the task of the researcher using the technology flow framework is to choose which technologies to highlight and which to abstract away. For example, display screens are essential to digital technologies, but they can safely be left out of most models because they do not figure in technology policy questions.

Third, questions of *who* makes the final combination of x_1 and x_2 and *where* they do it touch on the theory of the firm and become very important to the openness/control question. In some cases, the combination is made by the end user, in other cases in the facility of the supply or demand technology, and in some cases the end user must procure further services which may result in a multi-sided market situation.

4 The computer and the internet

Let us now turn to two specific examples. Rather like the steam engine and the railway, many lists of general purpose technologies include entries for both “the computer” and “the Internet.” Both may pass muster as GPTs, but it is much more common to refer to the Internet as “infrastructure” than the computer. We find this understandable through the technology flow framework: while both are widely used on the demand side, the facilities to supply the Internet are much less rivalrous in consumption than those that supply the computer.

Let us apply the framework to the traditional personal computer. The idea behind it could be termed “electronic computing.” It is the set of scientific principles and engineering know-how that allow a computer to be designed and factories to be

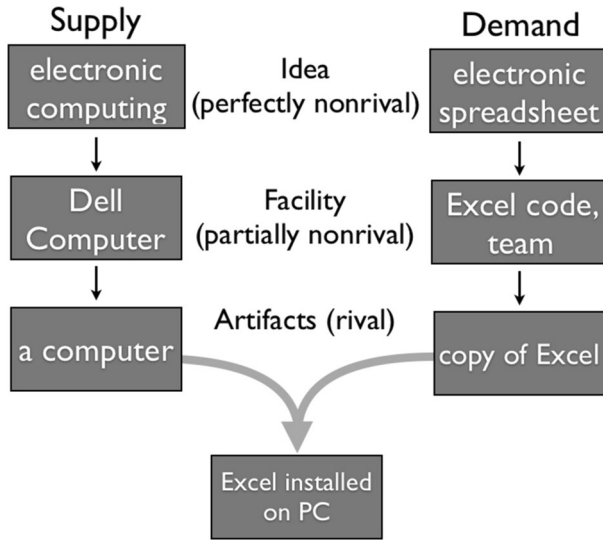


Fig. 3 The computer and one traditional use, the spreadsheet

built that can make computers.⁵ The facility to make computers is a factory. More accurately, it is a bundle of facilities that include not only factories but the design, marketing, service, and other functions that are usually bundled as a firm. Dell Computer is an example of one such facility organized as a firm. Ultimately, the output created by Dell and other computer makers is an artifact, a single computer that is a rival, private good.

An essential characteristic of computer production is that there is some rivalry. As manufactured goods, individual computers incorporate discrete chunks of materials, energy, labor time, and space. This means the facilities to create them use production functions with relatively high marginal costs. Computers also incorporate high fixed costs of development and support. The final good embodies a share of nonrival inputs as well. We could thus say that the computer production function exhibits partial nonrivalry, but not to a great extent.⁶

For the most part, users do not want a computer as a final consumption good. Rather, there are many different uses for the computer. One traditional use is the spreadsheet application, the original “killer app.”⁷ Again, this is based on a nonrival idea—the virtualization and automation of older paper methods of business analysis. The facilities to make a spreadsheet application generally involve a team of software developers and some source code—the Excel team from Microsoft is one example.

⁵ As always, one can further subdivide, and say that computers are built on microprocessors, storage devices, and so forth. Older GPTs like electricity and even the wheel (on the mouse or disk drive) play a role in computers.

⁶ An even more pure example of rival production would be Adam Smith’s pin factory, where the user of a pin would not need any design or support services at all.

⁷ Apparently the first reference to this term is Ramsdell (1980) regarding the *VisiCalc spreadsheet*.

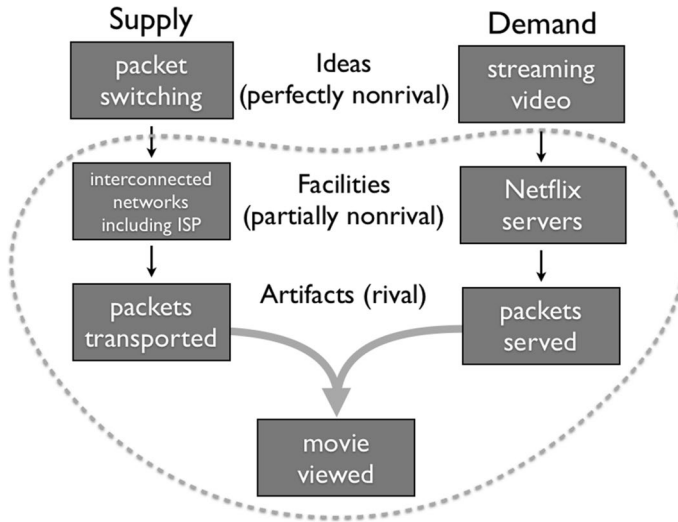


Fig. 4 The Internet and one use, Netflix video

The output of this team is a copy of Excel, another artifact. The two artifacts are then combined to create a working, installed spreadsheet application.⁸ This example is illustrated in Fig. 3.

In the traditional desktop environment, the end user combines the computer artifact and Excel disk artifact at his or her own premise. Neither Dell nor Microsoft are involved in this final installation. This allow for some freedom to the user, who might, for example, operate the computer using Linux and install Excel into a virtual machine rather than the standard situation where Windows is the computer's operating system.

Now we turn to a technology closely related to the computer, the Internet. The idea behind the Internet is packet-switched interconnection of computers. This idea can be implemented in many ways including local area networks within one building. But the most effective implementation has been an enormous facility comprised of interconnected networks including Internet backbones and local ISPs. This facility can produce a very potent artifact, a packet of data sent between computers around the world.

Internet packets are produced under partially nonrival conditions. There is very little marginal cost except when peak traffic causes congestion. But the only way to provide these scale economies is for the user to maintain a direct, ongoing connection to the facility that transmits the packets.

From the demand side, there are many uses of the Internet. One use often associated with the network neutrality debate is streaming video, which not only requires

⁸ The adjective *working* is important here. One could have a non-working copy of Excel even without a computer, just as one could have a non-working car without an internal combustion engine.

an ongoing connection to partially nonrival packet transport, but also an ongoing connection to the partially nonrival server capability of a company like Netflix. These connections imply a business relationship between the video streaming company and the packet transport service and thus invite additional business issues, such as pricing using two-sided market principles and disputes over discrimination. The combined artifact is an actual segment of video viewed by a user, as illustrated in Fig. 4. The dashed line in the figure represents the requirement of ongoing connection between the facilities and artifacts.

In both the spreadsheet and Netflix examples, the technologies can be broken down using the technology flow framework. The main difference is that in the Netflix example the artifacts cannot be combined into useful output without ongoing connection to the facilities. In both examples, there are alternative technologies that change the connection requirements. Online applications like Google Sheets make the spreadsheet into a packetized service that is quite similar to a Netflix movie. On the other hand, downloadable video creates a video file artifact that can function separately from the facility that created it and separately from the Internet.

5 Infrastructure and GPTs: similar on demand side, different on supply side

So what makes infrastructure different from other implementations of a GPT? Our answer is that on the demand-side, there is not much difference. Both are inputs into a wide range of downstream uses, as their definitions in Frischmann (2012) and Bekar et al. (2018) make clear. In both cases, the artifacts supplied are then used in a tremendous variety of applications that create some other kind of useful artifact. For example, both the internal combustion engine and electricity produce artifacts (engines and kilowatt hours) that are used in a huge number of downstream technologies to create other useful artifacts. The fact that it seems more natural to call electricity “infrastructure” than the internal combustion engine does not have much to do with the variety of uses available.

The difference lies with the supply-side facility. For many technologies, the facility is a factory. Some inputs used in a factory, such as designs, organization, and large-scale machinery are nonrival or partially nonrival. Others, such as labor time, materials, and energy are completely rival. The result is that supplying an additional artifact consumes real resources and has a significant marginal cost. In contrast, other technologies have less modular facilities with extremely large economies of scale; often they are networks built to connect many users. The material, energy, and labor inputs to these large-scale facilities are negligible for producing off-peak artifacts. Only when output becomes so high that the facility becomes congested does supplying an additional artifact create a significant marginal cost.

When an implementation of a GPT involves a facility with a scale of partially nonrival production that is large relative to the extent of the market, that use of the GPT can be termed *infrastructure*. The large size relative to the market may cause market imperfections in provision of the facility. There may be barriers to entry,

natural monopoly, and large difference between average and marginal costs—all problems familiar in regulatory economics.

To be sure, there are facilities that produce artifacts under conditions of partial nonrivalry but do not have a wide range of uses in the economy. Software code like Excel is an example, as are print newspaper printing plants and public or club swimming pools. The same might be said of a flour mill in a medieval town or a server farm for cloud computing. Sometimes facilities like these are casually referred to as “infrastructure,” but this use of the term is imprecise. Perhaps to be complete one should always differentiate between “special purpose infrastructure” and “general purpose infrastructure,” but we believe that the best use of the word really only refers to the general purpose case. When people call an implementation of a special-purpose technology “infrastructure,” the label is really intended to mean “important” or “essential to the activity.”⁹ The word “platform” is often used for implementations of technologies that are intermediate between highly specialized and very general, and this is probably more precise. Plantin et al. (2016) discuss the differences between a platform and infrastructure, but they note that digital technologies have allowed for “a ‘platformization’ of infrastructure and an ‘infrastructuralization’ of platforms” (p. 3).

Nonrival use of the supply-side facility is the key differentiator that makes the Internet an infrastructure as well as an implementation of a general purpose technology. In contrast, “the computer” is produced in a rival manner like any other manufactured good. So while the computer is a general purpose technology based on its demand-side uses, there is no implementation of the computer that can properly be called infrastructure. This says nothing about the relative impact or importance of the Internet versus the computer, only that the facilities that embody these technologies have different production properties.

The extent of nonrivalry is greater for infrastructure than for other uses of GPTs. For both, there are nonrival ideas for *use* that generate positive spillovers for the economy. But in infrastructure, there is also substantial nonrivalry in the supply-side *facilities*, and this creates important differences in how infrastructure is provided and regulated.

6 Openness and tethering

Demand-side uses for both the computer and the Internet are based on nonrival ideas. Like most ideas, they are open and nonexcludable except to the extent that they can be protected by intellectual property, such as trade secrecy or a patent. The result for our examples from Sect. 4 is that there are many competing sources of spreadsheet programs and of streaming video content, each a separate company with

⁹ In most cases that we are aware of, the owners of a facility seem to perceive benefits from being labeled infrastructure, since this implies that the facility is more worthy of attention and perhaps favorable policy treatment through regulation or even subsidies. However, if the label “infrastructure” is seen as implying social or regulatory obligations (more like “public utility”), facility owners might resist it.

its own facility. It is essential that each demand-side use be able to obtain access to the supply-side artifact, and this depends on whether there is a “manager” with control over access to these artifacts.

These examples show that the source of spillovers to the general economy is the openness of the ideas for using the technology and the openness of access to the supply-side artifacts. By having many ideas and allowing each idea to be reused by many agents, the range of uses of the technology rises and its value to the overall economy multiplies. By having the supply-side artifacts be open to every potential use, the various ideas for uses can be implemented by building the appropriate facilities and producing the necessary artifacts.¹⁰

Infrastructure facilities usually have huge economies of scale that make marginal cost negligible when congestion is not taking place. In most cases, the way such large economies of scale exist is that infrastructure operates as a network connecting various users. Sometimes this is explicit, as in the road and telephone networks that physically terminate at driveways and household jacks. Sometimes it is more virtual, as in seaports and wireless telecommunications that provide connections using the natural resources of the oceans and the electromagnetic spectrum. In either case, networking is an important feature of infrastructure that may not be present when GPTs are implemented through stand-alone components such as the steam engine or computer.

In cases where artifacts can only be obtained via an ongoing connection to the supply facility, the service is *tethered*. Zittrain (2008) defines tethered services as those which are “centrally controlled.” Tethering often relies on surveillance and control technologies over networked facilities. This is crucially important because it creates an ongoing relationship, often with legal ramifications, between the supply-side facilities provider and the demand-side user. Tethering is different from the notion of an *excludable* good (Ostrom and Ostrom 1977). Both tethering and exclusion, for example via property rights, facilitate appropriation by suppliers. But they raise different economic and policy issues because they work quite differently. Exclusion usually functions at the point of purchase of an artifact and enables a market transaction. Tethering functions on a continuous basis (post-purchase) and enables suppliers to execute different pricing and control strategies, including some that differentiate based on use.

In a stand-alone GPT, like the internal combustion engine, the maker of the engine owns the engine until it is sold, and thus it is excludable, but the engine-maker would find it difficult to prevent the engine’s use in, say, a tractor instead of a truck. Likewise, makers of computers cannot easily prevent them from being used for, say, electronic banking instead of word processing. In contrast, with a tethered service, the connected nature makes usage control much more feasible. For example,

¹⁰ There is a downside to this openness, since it enables uses that may produce negative third-party effects that are both legal (transportation of cigarettes or transmission of pornography) or illegal (transportation of illegal weapons or hacking of personal information). These negative third-party effects can lead both to call for policing while keeping the system open (the usual response with roads) or to calls for the system to be made less open (e.g. requiring Facebook to detect factually incorrect information).

if the telephone provider does not want the user to contact a dial-up ISP, it can simply block such service unless there is regulation to the contrary. Another example is a toll highway: these are usually required to serve all vehicles, but without regulation a toll road could easily discriminate against trucks of a certain company. Indeed, technology providers may purposely try to tether their facilities to their artifacts—Zittrain calls this “appliancizing” and Plantin et al. (2016) call it “platformization” of infrastructure. Odlyzko (2003, 2004) has shown that industries generally drive toward price discrimination when feasible, and modern tethering by digital networked technologies follows that trend. Policies to reduce negative third-party effects can sometimes create or strengthen tethering; this is especially prevalent in calls for social media platforms to police their networks for hate speech, factually-incorrect information, bot accounts, and other forms of undesirable content.

Liebowitz and Margolis (1994, pp. 135–136) noted this difference and used the term “literal networks” to describe what we call tethering: “In some networks, participants are literally connected to each other in some fashion. The telephone system is one such network, as are pipeline, telex, electrical, and cable television systems. These ‘literal networks’ require an investment of capital, and there is a physical manifestation of the network in the form of pipelines, cables, transmitters, and so on. It is not only feasible but almost inevitable for property rights to be established for these types of networks. Those who attach to such networks without permission from the owner, or who attach without adhering to the rules, may be disconnected, a characteristic that removes the problem of nonexclusion.”

We subdivide tethering into three categories:

1. *physical tethering*—the facility physically connects with those who use the artifacts (broadband networks, road networks).
2. *virtual tethering*—the facility requires (or is engineered to require) a non-physical connection with those who use the artifacts (wireless phones, iOS App Store).
3. *legal tethering*—a license or other legal permission is required to use the facility and obtain artifacts (patent licenses, software licenses, permits for use of public lands).

Because of the differences in the prevalence of tethering, infrastructure is usually “managed” by its provider, while implementations of GPTs in general may or may not be. Theories of GPTs emphasize the interdependence of the GPT itself and the many components that work with it. The need to create and coordinate many components may lead to prisoner’s dilemma games in which component makers free-ride and do not produce enough components and to coordination games where component makers need to agree on standards and interfaces. Even when there are no market failures of

the traditional economics type, the range of *technological complementarities* (Carlaw and Lipsey 2002) is very large for a GPT, and so any forces that block productive uses over time could have a large, negative effect on the economy.

GPTs like the steam engine or biotech experience untethered diffusion, and it is difficult to imagine “managing” them or preventing “permissionless innovation.”^{11,12} The tethered diffusion of GPTs implemented through infrastructure, like railroads or the Internet, makes them much more “manageable.” In particular, tethering means that two of Ostrom’s (1990) “design principles” for commons management apply more easily to infrastructure than to GPTs: clearly defined boundaries and monitoring.

Infrastructure has more “clearly defined boundaries” than other implementations of GPTs. Because infrastructure is tethered, it is more clear who is and is not using the infrastructure and how. Generally, those who wish to use infrastructure must formally become customers of the infrastructure provider or go through some related process such as registering their car in order to drive on public roads. The owner of a tethered technology retains “bouncer’s rights” (Strahilevitz 2005) to prevent use. Boudreau and Hagiu (2009, p. 7) note this and then add, “The power to exclude also naturally implies the power to set the terms of access (e.g. through licensing agreements)—and thus to play a role somewhat analogous to the public regulator.” But without tethering and its attendant power to monitor and exercise control over uses and users, the transactions costs of such control are usually overwhelming.

7 Implications

Policies regarding infrastructures are complicated because of the many linkages between an infrastructure and other sectors of the economy (Hogendorn 2012). Aghion et al. (2009) discuss the difficulty of linking theoretical approaches to real-world policies that promote science, technology, innovation, and growth systems (STIGS). They note the “GPT rationale” for public investment, but also caution that policy responses are more difficult when there are many complex complementarities rather than isolated market failures. Bauer (2014, p. 671) emphasizes that it is incumbent upon the “analyst to take the relevant interdependencies among players connected by a platform into account.” Our technology flow framework puts structure on this: infrastructure policy that affects *artifacts* only affects downstream factor prices, whereas policy that affects the *facility* may actually change downstream production functions or lead to the exit or entry of goods.

When an economics researcher writes a production function in either theoretical or empirical work, the technology flow framework can provide guidance. The

¹¹ This term seems to have been first used by Vint Cerf, David Reed, Stephen Crocker, Lauren Weinstein and Daniel Lynch in an Oct. 2009 letter to then FCC Chairman Julius Genachowski supporting network neutrality: <http://voices.washingtonpost.com/posttech/NetPioneersLettertoChairmanGenachowskiOct09.pdf>.

¹² Increasingly there are field-of-use restrictions in licensing agreements that legally “tether” the technology to certain uses (Schuett 2012).

production function should have (at least) two factors of production, x_1 and x_2 which are the artifacts produced by the supply-side and demand side facilities. Any capital is the capital that goes into the facilities themselves, say K_1 and K_2 .

Tethering is the reason why the openness/control question tends to revolve around infrastructure but not other types of general purpose technologies. Openness to innovation is equally important for all of them, but tethering creates a situation where innovative use often requires permission from infrastructure facilities providers. Beyond permission, infrastructure facilities providers can discriminate in various ways (e.g., price, priority, quality of service) to appropriate a greater portion of the surplus generated by innovative uses.

A great deal of research has tried to measure the impact of ICT on the economy. It generally follows on Paul David's seminal "The Dynamo and the Computer" (1990) which gives two GPTs in its title and keeps them parallel by choosing the device that serves as an engine rather than any system that goes with it. Jovanovic and Rousseau (2005) are also clear that by "IT" they mean the microprocessor and personal computer. But most recent papers expand from the computer to the whole category of ICT—see the survey by Cardona et al. (2013).

While the contribution of the ICT sector to the economy is a worthy question, "ICT" cannot be cinched into the technology flow framework. There is no single artifact which can be considered the output of ICT, and it would be difficult to identify a single group of facilities that enables all of ICT. This suggests that the computer and the Internet are two separate GPTs, just as in Lipsey et al.'s (2005) list.¹³

Empirical work on GPTs frequently searches for evidence of spillovers from the adoption of a GPT, either in R&D productivity or in total factor productivity (TFP) growth. Some authors use the strength of spillovers to *test* whether a technology can be labelled a GPT, for example Feldman and Yoon (2011) and Liao et al. (2016). Others use the definition to declare a technology a GPT prior to further analysis, for example Jovanovic and Rousseau (2005). We believe our framework helps with either approach, either to define a GPT or to select a candidate GPT that will be subject to empirical testing. But in both approaches, it suggests that spillovers should be measured across supply–demand sectors that can be modeled using the technology flow framework. If this is taken to mean that ICT should be separated into computers and the Internet (or possibly other sectors such as content), this would reframe the question of spillovers across sectors.

The most important part of this reframing would be to change the view of spillovers that occur *within* the category ICT as defined by statistical authorities. Consider that computers and the Internet are both categorized in this broad sector of the economy, whereas steam engines and railroads are categorized as separate sectors. Yet the contribution of computers to the Internet seems entirely comparable to the contribution of the steam engine to railroads.

¹³ Carlaw et al. (2008) present a different view that there is one overall GPT called "programmable computing networks" of which computers and the Internet are just separate implementations. They also argue that the term ICT should properly be even broader, encompassing writing systems, printing, and so forth since these also facilitate communication of information.

This more nuanced view alters the research questions about spillovers within ICT. For example, imagine a country that did not produce computers but had a healthy Internet sector based on imported computers. This would depend very much on the degree of tethering and the permissiveness of the innovation environment. The move from the general purpose PC to the more appliancized phone and tablet could affect the hypothetical country's tech sector. Venturini (2015) found that OECD countries received high TFP spillovers from domestic production of ICT goods but not from importing ICT goods. Perhaps this is because imported technology remains tethered to foreign firms and thus cannot serve as an innovation platform to create downstream spillovers in the importing country.

As the combination of artificial intelligence and big data becomes more important, it too may be considered a GPT. The technology flow framework would suggest that it too should be considered a separate generic technology, though it may sometimes be usefully placed under the umbrella term ICT. An important question will be whether artificial intelligence and big data are implemented as infrastructure and involve tethered diffusion under centralized control. Currently tethered diffusion seems like the most likely implementation strategy, in which case these technologies are likely to result in a new form of infrastructure along with the attendant questions of openness and control.

8 Conclusion

We have discussed two important terms, general purpose technology (GPT) and infrastructure. The exact relationship has not been clear in the literature. Here we argue that the key is to split up implementations of technology into their demand-side and supply-side properties. On the demand side, all GPTs have many demand-side uses by definition, but on the supply side GPTs may be implemented through facilities that have varying levels of partial nonrivalry. When the scale of partially nonrival production is large relative to the extent of the market, the implementation of the GPT can be labeled infrastructure and is more likely to be subject to regulation and also more likely to employ commons management techniques. In particular, infrastructure is usually built as a network and involves some *tethering* of the user to the provider, which means that maintaining boundaries and monitoring use is much easier in the context of infrastructure. This in turn explains why the openness principle becomes more important in the infrastructure context.

We hope our technology flow framework will be useful in two ways. First, it provides guidance on how and where to look for positive spillovers that are often relevant in regulatory and antitrust proceedings. Second, it clarifies the use of the terms infrastructure and general purpose technology which are themselves often used to bolster contentious claims in policy debates. We also hope that it will help spur further research on infrastructure as a microfoundation of economic growth.

References

- Aghion, P., David, P. A., & Foray, D. (2009). Science, technology and innovation for economic growth: Linking policy research and practice in 'stig systems'. *Research Policy*, 38(4), 681–693.
- Arthur, W. B. (2009). *The nature of technology: What it is and how it evolves*. New York: Free Press.
- Aschauer, D. A. (2000). Public capital and economic growth: Issues of quantity, finance, and efficiency. *Economic Development and Cultural Change*, 48(2), 391–406.
- Bauer, J. M. (2014). Platforms, systems competition, and innovation: Reassessing the foundations of communications policy. *Telecommunications Policy*, 38(8–9), 662–673.
- Bekar, C., Carlaw, K., & Lipsey, R. (2018). General purpose technologies in theory, application and controversy: A review. *Journal of Evolutionary Economics*, 28, 1005–1033.
- Boudreau, K. J., & Hagiu, A. (2009). *Platform rules: Multi-sided platforms as regulators*. London: Edward Elgar Publishing.
- Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies 'engines of growth'? *Journal of Econometrics*, 65(1), 83–108.
- Cardona, M., Kretschmer, T., & Strobel, T. (2013). ICT and productivity: Conclusions from the empirical literature. *Information Economics and Policy*, 25(3), 109–125.
- Carlaw, K. I., & Lipsey, R. G. (2002). Externalities, technological complementarities and sustained economic growth. *Research Policy*, 31(8), 1305–1315.
- Carlaw, K. I., Lipsey, R. G., & Webb, R. (2008). Has the ICT revolution run its course? *International Congress on Environmental Modeling and Software*, 262, 200–212.
- David, P. A. (1990). The dynamo and the computer: An historical perspective on the modern productivity paradox. *The American Economic Review*, 80(2), 355–361.
- David, P. A., & Foray, D. (1996). Information distribution and the growth of economically valuable knowledge: A rationale for technological infrastructure policies. In M. Teubal, D. Foray, M. Justman, & E. Zuscovitch (Eds.), *Technological infrastructure policy* (pp. 87–116). Berlin: Springer.
- FCC. (2010). *2010 open internet order*. FCC 10-201, December 23, 2010.
- FCC. (2015). *2015 open internet order*. FCC 15-24, March 12, 2015.
- Feldman, M. P., & Yoon, J. W. (2011). An empirical test for general purpose technology: An examination of the cohen–boyer rdna technology. *Industrial and Corporate Change*, 21(2), 249–275.
- Field, A. J. (2011). *A great leap forward: 1930s depression and US economic growth*. New Haven, CT: Yale University Press.
- Frischmann, B. M. (2012). *Infrastructure: The social value of shared resources*. Oxford: Oxford University Press.
- Frischmann, B. M., & Hogendorn, C. (2015). Retrospectives: The marginal cost controversy. *Journal of Economic Perspectives*, 29(1), 193–206.
- Frischmann, B. M., & Lemley, M. A. (2007). Spillovers. *Columbia Law Review*, 107(1), 257–301.
- Frischmann, B., & Selinger, E. (2018). *Re-engineering humanity*. Cambridge: Cambridge University Press.
- Ghosh, A., & Meagher, K. (2004). *Political economy of infrastructure investment: A spatial approach*. North American Econometric Society Summer Meetings, Brown University.
- Gramlich, E. M. (1994). Infrastructure investment: A review essay. *Journal of Economic Literature*, 32(3), 1176–1196.
- Grimes, A. (2014). Infrastructure and regional economic growth. *Handbook of Regional Science*, 331–352.
- Helpman, E. (1998). Introduction. In E. Helpman (Ed.), *General purpose technologies and economic growth*, Cambridge, MA: MIT Press.
- Helpman, E., & Trajtenberg, M. (1998). Diffusion of general purpose technologies. In E. Helpman (Ed.), *General purpose technologies and economic growth* (pp. 85–119). Cambridge, MA: MIT Press.
- Hess, C., & Ostrom, E. (2003). Ideas, artifacts, and facilities: Information as a common-pool resource. *Law and Contemporary Problems*, 66(1/2), 111–145.
- Hogendorn, C. (2005). Regulating vertical integration in broadband: Open access versus common carriage. *Review of Network Economics*, 4(1), 19–32.
- Hogendorn, C. (2012). Spillovers and network neutrality. In G. Faulhaber, G. Madden, & J. Petchey (Eds.), *Regulation and the performance of communication and information networks* (pp. 191–208). London: Edward Elgar Publishing.

- Jovanovic, B., & Rousseau, P. L. (2005). General purpose technologies. *Handbook of Economic Growth*, 1, 1181–1224.
- Justman, M., & Teubal, M. (1995). Technological infrastructure policy (tip): Creating capabilities and building markets. *Research Policy*, 24(2), 259–281.
- Liao, H., Wang, B., Li, B., & Weyman-Jones, T. (2016). Ict as a general-purpose technology: The productivity of ICT in the United States revisited. *Information Economics and Policy*, 36, 10–25.
- Liebowitz, S. J., & Margolis, S. E. (1994). Network externality: An uncommon tragedy. *Journal of Economic Perspectives*, 8(2), 133–150.
- Lipsey, R., Carlaw, K., & Bekar, C. (2005). *General purpose technologies and long-term economic growth*. Oxford: Oxford University Press.
- Murphy, K. M., Shleifer, A., & Vishny, R. W. (1989). Industrialization and the big push. *Journal of Political Economy*, 97(5), 1003–1026.
- National Research Council. (1987). *Infrastructure for the 21st century: Framework for a research agenda*. Washington, DC: National Academy Press.
- Odlyzko, A. (2003). Privacy, economics, and price discrimination on the internet. In *Proceedings of the 5th international conference on electronic commerce* (pp. 355–366), ACM.
- Odlyzko, A. (2004). The evolution of price discrimination in transportation and its implications for the internet. *Review of Network Economics*, 3(3), 323–346.
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge: Cambridge University Press.
- Ostrom, V., & Ostrom, E. (1977). Public goods and public choices. In E. S. Savas (Ed.), *Alternatives for delivering public services: Toward improved performance* (pp. 7–49). Boulder, CO: Westview Press.
- Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2016). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, 20, 293–310.
- Ramsdell, R. E. (1980). The power of VisiCalc. *BYTE*, November, 190–192.
- Rosenberg, N. (1978). *Inside the black box: Technology and economics*. Cambridge: Cambridge University Press.
- Rosenstein-Rodan, P. N. (1943). Problems of industrialisation of Eastern and South-Eastern Europe. *The Economic Journal*, 53(210/211), 202–211.
- Schuett, F. (2012). Field-of-use restrictions in licensing agreements. *International Journal of Industrial Organization*, 30(5), 403–416.
- Steinmueller, W. E. (1996). Technological infrastructure in information technology industries. In M. Teubal, D. Foray, M. Justman, & E. Zuscovitch (Eds.), *Technological infrastructure policy: An international perspective* (pp. 117–139). Berlin: Springer.
- Strahilevitz, L. J. (2005). Information asymmetries and the rights to exclude. *Michigan Law Review*, 104, 1835.
- Venturini, F. (2015). The modern drivers of productivity. *Research Policy*, 44(2), 357–369.
- Zittrain, J. (2008). *The future of the internet—And how to stop it*. New York: Yale University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.