

The Production of Increasing Returns

Physical Technology, Institutional Technology, and the Pitfalls of Production Functions

Santiago Gangotena*

Ryan Safner†

June 24, 2022

Abstract

Economists often model the delicate relationship between production, innovation, and economic growth as a production function that exhibits increasing returns to scale (IRS). The existence of knowledge spillovers or learning by doing implies conclusions about the optimal use of protectionist trade and industrial policies that increase national welfare by reallocating productive resources in order to exploit IRS. We claim that all physical technologies, understood as recipes that produce a specified output from given inputs, exhibit constant returns to scale. Increasing returns to scale comes from institutions that enable entrepreneurial discovery and choice to *switch* between physical technologies. We examine several case studies of apparently successful infant industries and clarify the nature and economic relationships behind the phenomena of increasing returns and the limitations of existing models of production, we contribute a new understanding of research with applications to innovation, economic growth, and trade policy.

JEL Classification: D20, O25, O38

Keywords: Technology, innovation, increasing returns, learning by doing, production functions, protectionism, industrial policy, institutions

*Universidad San Francisco de Quito USFQ, Colegio de Administración y Economía; sjgangotena@usfq.edu.ec

†Hood College, Department of Economics and Business Administration; safner@hood.edu

1 Introduction

Theoretical justifications for both industrial policy and trade restrictions can best be characterized as variations on a theme: in the presence of increasing returns to scale (IRS) there are potential net gains from restraining free trade, limiting competition, or changing the structure of production.¹ Although IRS as a justification for protectionist measures fell out of academic favor for some time, it has recently resurfaced under the guise of “knowledge spillovers” stemming from the success of endogenous growth models (Romer 1986, 1990). Some modern international trade researchers argue for new industrial policy programs aiming to exploit IRS in the creation of knowledge (e.g. Matsuyama (1992), Rodriguez and Rodrik (2001), Rodrik (2004a), Rodrik and Subramanian (2005), Greenwald and Stiglitz (2006), Hausmann, Hwang, and Rodrik (2007), Stiglitz and Greenwald (2014)). The analytic framework and policy stance may be summed up by Stiglitz (2014):

Industrial policies – in which governments intervene in the allocation of resources among sectors or favor some technologies over others – can help “infant economies” learn. Learning may be more marked in some sectors (such as industrial manufacturing) than in others, and the benefits of that learning, including the institutional development required for success, may spill over to other economic activities.

We contend, however, that present modeling techniques and their policy implications often lack considerations about the prerequisite *institutional* conditions necessary for increasing returns to exist in the first place. We fear policy recommendations derived from models that fail to specify exactly how IRS emerges risk destroying the very advantage those policies are intended to exploit.

A production process exhibits increasing returns when increasing all inputs by some amount results in a more-than-proportionate increase in output.² For analytical tractability, since Marshall (1920), economists have modeled IRS by via a production function with a firm’s output increasing with the scale of the entire industry’s output. The intuition is that as aggregate industry output increases, individual firms become more efficient and adept at using more advanced production technologies. In certain industries, knowledge is often produced as a byproduct of the production process itself (e.g. information technology, pharmaceuticals, or manufacturing) or where firms explicitly invest in the creation of new knowledge through research & development,³ or where producers discover methods of tweaking production routines to enhance efficiency

¹Historically, economists have variously referred to this phenomenon as “economies of scale” or “external economies” (Marshall 1920), “decreasing costs” (Graham 1923; Knight 1924, 1925), and “increasing returns” (Romer 1986). For convenience and consistency, we use “increasing returns to scale” (IRS).

²Formally, $f(cK, cL) > cf(K, L)$ where $f(cK, cL)$ is a production function utilizing inputs capital (K) and labor (L), and $c > 0$.

³The production of a non-rival good such as knowledge features high fixed costs of research and development, but any

simply by continuous repetition and/or serendipitous discovery (“learning by doing”). Nonrival knowledge arising from any of these proposed mechanisms can spill over from one firm to be exploited by all firms in the industry.

These arguments build on the classic economic debate about optimal tariffs, as well as a secondary argument about the incompatibility of perfect competition with industries with increasing returns. The classical economists seemed to take it as an empirical law that agriculture exhibits systematic decreasing returns to scale while manufacturing exhibits systematic increasing returns to scale.⁴ Marshall (1920, bk. V, Chapter XII) echoes this approach, proposing that a subsidy to an industry exhibiting IRS could provide a net benefit to consumer-taxpayers. The neoclassical theories of perfect competition and marginal productivity theory of distribution hold that in competitive (factor) markets, all factors of production are paid according to the marginal contribution they make towards the value of the final product. Clark (1889, 1891) famously contended that when each factor is paid its marginal product, adding up all of the marginal productivities “fully exhausts the product,” and Wicksteed (1894) proved this assertion by using Euler’s Theorem by *assuming* constant returns to scale.⁵ Much of this debate took on a secondary moral dimension regarding the apparent natural justice that under the product exhaustion theorem, competitive markets deliver to each factor their just deserts. Graham (1923) makes perhaps the first sophisticated case that countries can increase net aggregate welfare by levying protective tariffs on industries with IRS and subsidizing the allocation of labor towards them, even at the expense of specialization in industries where there is an initial comparative advantage.⁶ Knight (1924) counters with several insights that strike at the root of the issue. First, the difficulty of identifying IRS: not only is IRS for individual firms incompatible with perfect competition,⁷ but more fundamentally that what appear to be external economies for *one* firm are internal and exploitable for *other* firms. Knight points out that because most production inputs are rivalrous, new entrants or increased output will bid costs upward, rather than lower them and concludes (332),

it must be shown that there are, or may be, industries, in a condition of stable competition, in

resultant discoveries can be *reproduced* or copied by others at a very low marginal cost, meaning the firm’s average cost often decreases with higher levels of (industry) output.

⁴“The most opulent nations, indeed, generally excel all their neighbours in agriculture as well as in manufactures; but they are commonly more distinguished by their superiority in the latter than in the former,” Smith (1776, Book I, Ch. 1).

⁵Consider an aggregate production function $Y = f(L, K, T)$, where Y is output, L is labor, K is capital, and T is land. If and only if this production function is linear and homogenous of degree 1, that is, for $cY = f(cL, cK, cT)$ and $c = 1$, then by Euler’s theorem, it can be written as $Y = MP_L L + MP_K K + MP_T T$, where MP_i represents the marginal product of factor i (which can be expressed as $\frac{\partial Y}{\partial i}$). This means the production function exhibits constant returns to scale and the total value of the product is exhausted by the marginal productivity of each factor. Note that this formulation further requires that each factor of production is continuous, rather than discrete marginal units (Machlup 1937).

⁶Taking the classic comparative advantage story between two countries and two sectors, if the production of one sectors (e.g. agriculture) exhibits decreasing returns, and the other sectors (e.g. manufacturing) exhibits increasing returns, then the more the two countries specialize in one sectors, the relatively cheaper both goods become for the country specializing in manufacturing, and the relatively more expensive both goods become for the country specializing in agriculture.

⁷“If competition is effective, the size of the productive unit will tend to grow until either no further [technical external] economies are obtainable, or there is only one establishment left and the industry is a monopoly,” Knight (1924, 597).

which no producer already engaged could decrease his real costs by expanding his output at the expense of other producers, and yet in which real costs would be decreased all around by new producers entering the industry in competition with those already there.

Modern arguments for trade restrictions rest upon economists and policymakers locating a realistic source of systematic IRS for all firms in an industry. Endogenous growth theory models propose that “knowledge” — when viewed as a non-rivalrous input — is a major source, as it can decrease costs for all individual firms in proportion to the scale of the industry’s output (see e.g. Arrow (1962b); Romer (1986); Romer (1990)). In the process, these models abandon the traditional benchmark assumptions of perfect competition in favor of imperfect competition, in order to be compatible with IRS (Krugman 1979, 2009; Arthur 1989, 1994, 1996). In these models, an increase in industry output caused by an increase in the employment of the rivalrous factors of production increases the available stock of knowledge, which is modeled as an increase in the productivity of all firms in the industry.⁸ In seeking to explain the wealth of nations, the rise of endogenous growth models have shifted the spotlight of explanatory power onto human capital and knowledge creation. “Knowledge” is a non-rival resource, since one person learning something does not preclude others from learning the same thing, and often stimulates the learning of others, who can benefit from using knowledge that “spills over” from the initial discoverer (Arrow 1962a). The non-rivalrous nature of knowledge enables increasing returns to the rivalrous factors of production (e.g. land, labor, capital) when producers increase the stock of knowledge.

A quintessential example of modern, IRS-based approaches to industrial policy is Rodriguez and Rodrik (2001, 269–72), which puts forth a simple model based on Matsuyama (1992). Their model postulates a country with two sectors, (1) agriculture, which features constant returns to scale, and (2) manufacturing which features increasing returns to scale, argued (as above) to exist due to knowledge spillovers, with labor as the sole (rivalrous) input for both industries. The authors postulate IRS in manufacturing by assuming the productivity of labor in manufacturing increases proportionally with total output of the entire manufacturing sector. Assuming the country’s initial comparative advantage is in agriculture, a properly calibrated tariff on manufactured imports which increases the profitability of domestic manufacturing will attract a greater share of labor into the domestic manufacturing industry. This will lead to greater manufacturing industry output, which in turn further increase manufacturing firms’ productivity, resulting in the virtuous cycle that increases *overall* domestic output. As their simulation results show, at some point, the dynamic gains in productivity are offset by the static efficiency loss of the trade tariff such that the optimal tariff $0 < \tau < 100\%$.

⁸Formally, let firm i ’s total factor productivity $a_i \propto A \ \forall i$, where A is industry-wide total factor productivity. Various models differ, but tend to define $A = f(n, k, h)$, that is, A is some function of population n , physical capital k and human capital h .

Hence, by properly identifying the source of IRS, industrial policies which restrict competition and free trade can potentially achieve greater economic prosperity and efficiency.

The troubling feature of these modeling techniques is that some economists have drawn normative conclusions of calling for industrial and trade policies to protect domestic industries where they suspect IRS are present, without specifying how policymakers are equipped to successfully identify and exploit increasing returns without jeopardizing the existence of those returns. By the familiar theoretical argument of market failure, knowledge production is said to be under-provided since the social value of the knowledge cannot be fully appropriated by the initial producer [Arrow 1962 appropriability]. To improve upon this suboptimal equilibrium, industrial policies which increase the employment of other rivalrous factors (such as labor) can potentially increase the total output of the industry, and hence firm productivity through a virtuous circle.

Economists also emphasize that the key role of institutions in for determining economic outcomes and policy feasibility sets (North 1990; Acemoglu, Johnson, and Robinson 2001, 2002; Glaeser et al. 2004; Rodrik 2004b). One vital, if neglected, institutional filter consists of enabling (or preventing) individuals, firms, and entire societies to capture the benefits from discovery and innovation in production techniques, allowing production at much larger scales and lower costs. As we will show below, none of the modeling techniques described above imply that IRS is a *fundamental* feature of the production of knowledge that can readily be exploited through policy. The critical assumption driving the results is *defining* productivity growth proportionately to total industry output, implying that IRS are *physically inherent* in certain production process, *independent* of any institutional regime. It is precisely this assumption that we seek to challenge by examining the use of the production function approach to modeling, and analyzing how it relates to real world economic activities and policy implications.

In pursuit of this goal, we advance two main claims: First, *technology*, properly understood as a recipe that physically relates inputs with a specified output can only exhibit *constant* returns to scale (CRS). Second, as a corollary, the existence of increasing returns arises from entrepreneurial *choices to discover and switch* between physical technologies, which is contingent on institutional environments where rivalrous competition plays a prominent role. Our critique can be viewed as a variant of the Lucas (1976) critique of policymaking that exploits modeled parameters as it applies to the parameters of the production function: Briefly restated, our claim is that the structural parameters of production functions that exhibit IRS are not “deep” inherent features of physical technology, but represent emergent and context-dependent features. Thus, in order for policy prescriptions based on IRS to obtain desired outcomes, one must first demonstrate that the institutional features that cultivate IRS will actually be present in the desired policy regime. After making the theoretical case for these two claims by differentiating between the contributions of “physical”

and “institutional” technologies to increasing returns, we apply these conclusions to specific instances of trade and industrial policies. By exploring the critical “intermediate step” of determining whether a production process will generate increasing returns under a specific institutional setting to yield the desired outcome, we can better understand when IRS is present, and consequently, when it is relevant for industrial and trade policy. Failure to heed this exploratory caution results in a fatal case of confusing an illustrative abstract model with the actual activities of people in real world markets.

We consider two case studies of historical uses of the “infant industries” argument that led to apparent success: tinplate manufacturing (Irwin 2000) and steel rail production (Head 1994) in the late 19th century United States. In both cases, the industries received tariff protection in order to achieve economies of scale to become competitive with initially-cheaper foreign competition, and protection was eventually removed. In both cases, the attainment of increasing returns were largely successful because of the *institutions* that enabled knowledge spillovers and learning-by-doing, primarily the unrestricted immigration of skilled workers from England and Wales. We conclude that while it is possible to retrospectively identify industries that successfully attained increasing returns with, or despite, policy interventions, it is much more difficult to identify *ex ante* industries that *inherently* (i.e. invariant to institutions) feature unexploited IRS.

2 Technology: Physical and Institutional

At the simplest level, the process of transforming inputs into output is commonly modeled in economics as a production function. The transformation is governed by some physical production technology, such that this technology is embodied in, or represented by, the production function. Consider the production of oral rehydration solution. This production process is governed by a formula that specifies fixed proportions of the ingredients (i.e. water, salt, and sugar), a list of tools (i.e. a measuring spoon and mixing container), and effort (labor or mechanical) required to produce the output (a fixed quantity of the solution). The relationships between the quantity of n inputs (x_1, x_2, \dots, x_n) and the quantity of output y can then be represented by a production function, $y = f(x_1, x_2, \dots, x_n)$. The technology represented by the production function can be thought of as a fully-specified recipe for the production of rehydration solution. It is important to note that “labor” as an input in the above production function serves a purely *mechanical* role of assembling the ingredients in the specified way. No economically-relevant capacities of labor, such as entrepreneurial awareness or individual choice, are necessary to obtain the precisely specified output. In this sense, the physical effort provided by human labor could, in principle, be substituted with a properly designed machine. In this example, the technology as captured by the production function merely relates the physical and chemical relationships between inputs and outputs, and there is no necessary economic role

for human agency, even though a human does exert labor effort.

Any real production process, however, is not a spontaneous physical phenomenon, it requires human agency, primarily embodied in the act of choice. Consider a single owner-manager who initiates a completely automated production process that will assemble the rehydration solution simply by pressing a button. Even in the simplest arrangement, human agency is required to judge whether it is worthwhile to push the button at all.

Of course, the situation that resource owners and production managers face is not captured by the simple decision to push a button, rather it is one of “rapid adaptation to changes in the particular circumstances of time and place,” (Hayek 1945, 524). At each moment in time, the owner-manager makes adjustments determined not only by physical and technological constraints, but also by her own choices and valuation of alternative courses of action in the face of her limited and imperfect knowledge. The complexity of the problem, and the relevance of human choice, grows exponentially when we further recognize that most production units (firms, industries, nations) consist of large multitudes of individual decision-makers, each with limited and often contradictory knowledge, differing purposes, and limited authority, mutually adjusting to one another in order to make their several plans come to consistent fruition. This collection of active choices by many individuals, or even that of the single owner-manager, is qualitatively different than the mere natural laws that govern the physical transformation of inputs to outputs. Yet, confusingly, *both* aspects of production are necessarily embedded in the production function when it is used to model real world production units.

At this point, it is apparent that the word “technology” as used in economic theory, captures the combination of two distinct processes arising from different sources, each with their own specific dynamics. To continue using the word “technology,” but to clarify its dual meanings, we propose to differentiate between (1) *physical* technology, and (2) *institutional* technology⁹. To the extent that a production function represents the *physical* transformation of inputs into output, “technology” (in the sense that economists typically use) refers to physical processes governed by natural scientific laws. The remainder of all other aspects of production is governed by “institutional technology,” where individual choices enter into social interactions that are embedded within specific institutions.¹⁰

Equivocation of the word “technology” between physical and institutional relationships often obscures the

⁹We eagerly await a more marketable term

¹⁰Others have observed and analyzed the existence of this level of “technology” by other names. Leeson and Boettke (2009) discuss a higher tier of entrepreneurship where private individuals discover and innovate new property-protecting institutions where state institutions fail to do so. Martin and Thomas (2013) apply this concept to politics, where political entrepreneurs seek to change constitutional rules when their ends are not attained through existing rules, using the U.S. Congressional committee system as an example. Safner (2016) applies it to studying how Wikipedia creates a governance structure facilitating the creation of a free online encyclopedia without financial incentives.

subtleties involved in production and in fostering national prosperity. It is easy to conflate the set of products that individuals could potentially produce in ideal institutional settings (e.g. perfect competition, efficient governments, and complete contracts) and the natural laws relating inputs and outputs. Failing to cautiously account for these differences could quickly lead to an unsupported conclusion. For example, one might incorrectly conclude that the *physical* relationships between computer inputs that make up the “recipe” for building a computer are the source of IRS, and therefore political authorities should implement regulations to e.g. subsidize, cartelize, or otherwise alter the institutional makeup of the computer industry to promote overall national prosperity.¹¹ This would sweep under the rug the critical role of existing market and regulatory institutions facilitating the ability of computer industry entrepreneurs to operate at large scales and to invest in complementary capital goods. Trade and industrial policies seeking to exploit IRS in computers might produce unintended consequences that weaken the institutions supporting increased scaling and investment, leading to a *less* productive computer industry. Such policies might, for example, subsidize inefficient methods and crowd out the entrepreneurial discovery of more efficient methods.

In order to carefully separate the physical and institutional processes present in any production unit, we find it necessary to pinpoint their respective determinants. In the case of physical processes, it is fairly straightforward: constant natural laws determine the processes of production (such as basic chemistry and physics determining the oral rehydration solution). Institutional processes, on the other hand, are determined by the decisions of many individuals both *within* and *outside* the production unit. The decisions of each individual involved in production are contingent on the incentives and information they face, which in turn are shaped by the institutions within which production takes place.

Within the production unit, there may be internal rules of conduct or general norms of business dealings between collaborating firms. These rules guide and constrain the decision power of each individual involved, and direct the flow of information, often hierarchically (Coase 1937; Williamson 1985).

External to any production unit, social institutions assist producers in making choices. Institutions such as money, market prices, property rights, and legal institutions that permit rivalrous competition, and private profits and losses, if robust and socially productive, assist individuals in economic calculation and prudent investment, and incentivize the production of goods that consumers value. Where these external institutions are weak, such as in the presence of open war, state expropriation, or lack of social trust, individuals will make very different production decisions, if any at all.

¹¹It is interesting to note that the often celebrated “Moore’s Law,” describing the consistent empirical observation that the number of transistors in an integrated circuit doubles approximately every two years, is a proposed *economic* law, rather than a technological one. One could easily foresee the relationship disappearing under various structural and regulatory changes in the high tech industry.

Across the literature, applications of production functions sometimes explicitly emphasize either the physical or the institutional phenomena in production. At one extreme, the so-called “replication argument” used to justify constant returns to scale in most economic textbooks (e.g. Barro and Sala-i-Martin (2004, 27)), which we explore in more detail below, hinges on the physical aspects of production processes. At the other extreme, Yoon and Buchanan (1999) justify the use of *generalized* IRS driven by increased division of labor and specialization, where both are social phenomena present in certain institutional settings. Of course, theoretical models can afford to displace one or the other feature to make a specific point. However, this can not be done when models are used to inform *policy* because *both* phenomena are always present in any real world production process. We next separate out the two aspects of “technology” as modeled by a production function to advance our two propositions:

Proposition 1: *Physical technology is a recipe for relating inputs to output, and only exhibits constant returns to scale (CRS).*

Proposition 2: *Increasing returns to scale (IRS) arises only from particular institutional technologies that enable the discovery of, and switching to, different physical technologies.*

2.1 Constant Returns: Physical Technology

Only the physical aspect of technology, as represented by a production function, is inherent to the physical process and is policy-invariant. No policy or institution can change the physical and chemical facts that water, sugar, and salt in specified amounts make oral rehydration solution. As such, the physical aspect of technology, as part of what a production function represents, can only exhibit constant returns to scale (CRS).

Economists often recognize CRS as in the physical aspect of production and justify this assumption in production functions by pointing out the ability to *replicate* a production process: any existing physical production process can be scaled larger simply by repeating the original process.¹² It is precisely in this sense that production technologies can correctly be thought of as “recipes” for transforming inputs into output. If one wants *twice* the amount of oral rehydration solution, simply *double* the amount of water, sugar, and salt or in other words, follow the same recipe *twice*. Due to the constancy and universality of natural laws that govern physical technology, by definition, any recipe will produce a specific quantity of output in every case where the specific quantities of inputs are present.

¹²A physical production process might only be feasible after a certain scale. Furthermore, for processes that use discrete inputs, it may only be possible to scale a process in integer amounts (Machlup 1937). These arguments, however, are not particularly relevant to our analysis.

The production function model is a limiting case of a fully-specified recipe. To any baker or chef, a recipe in colloquial terms includes both a list of ingredients, the output (each in specified quantities), and a series of steps for combining the ingredients, i.e. an algorithm. Economists ignore the algorithm of production and focus entirely on the ingredients and outputs. This brings parsimony to analysis, as algorithms are often contextual and not easily generalized, whereas the “ingredients list” approach is abstract yet widely applicable. In our view, a recipe of physical technology fully specifies the quantities of all inputs, the quantity of output, and the sequence of steps involved in the transformation. In other words, physical technology is the aspect of production that can be represented by a properly-specified algorithm, and hence can potentially be fully automated without any need for human labor. To bake one serving of cake, a recipe calls for so much flour, yeast, sugar, salt, etc. and a series of steps detailing how the inputs yield the cake. A well-designed machine could execute this process without any need for human participation in the baking process once the baker makes the decision to bake a cake. Furthermore, changing (the quantity of) any ingredient (or in some cases, changing the sequence of steps involved) necessarily implies a violation of the algorithm or recipe of the physical technology being employed, and further implies a switch to a *different* physical technology. Thus, we again contend that due to the ability to replicate any fully-specified recipe, all physical technologies can *only* exhibit constant returns to scale. As a corollary, increasing returns to scale can arise only from *institutional* technology.

A number of writers have admirably attempted to reestablish production theory without recourse to the “black box” of production functions taking a structural approach to production and institutions. Following the work of Babbage (1835) on machinery, Scazzieri (2014) aims to ground production on stronger microfoundations that may exhibit increasing returns may exist. He notes that “certain enabling conditions must be satisfied for increasing returns to be *feasible*...[yet] enabling conditions are not sufficient for increasing returns to be *achieved*,” (76, emphasis in original). Beyond Smith’s benefits of the division of labor — increases in dexterity, time-saving, and invention of machines — Scazzieri (2014, 77) introduces Babbage’s fourth benefit: a *structural* principle that splitting a complex production process into sub-systems and organizing them proportionately will maximize the utilization of inputs. Rather than indivisibilities of inputs, this highlights an indivisibility in the production process. Scazzieri (2014, 82) further separates increasing returns from technical progress, identifying the former as productive improvements that arise out of adapting existing ideas, and the latter as new inventions. Focusing on the *structural* elements that *enable* increasing returns (rather than the institutional elements that govern its actual dynamics), he builds a theory of increasing returns inherent in the physical nature of certain production processes. “As a result, increasing returns are explained by the compositional principles governing the relationships among elements of the production

process and by the opportunities arising from within existing production organizations,” (ibid, 82).

In our view, we see some of these structural elements not as part of a particular recipe (physical technology), but the process of switching between recipes (through subdividing production processes, organizing them within or across production units, etc) is described by what we are terming “institutional technology.” While we applaud such a reinterpretation of production theory without recourse to production functions, our approach differs primarily in focus and semantics. In our view, we describe entrepreneurial choice within various institutions that constrain, enable, and orient individuals (Cardinale 2018) towards various production recipes that allows the phenomena described by increasing returns to take place. We agree that some physical features of certain production processes may *enable* potential increasing returns within an industry — some produced goods may benefit more than others from more roundabout production methods, substitutes in materials, or complementary capital goods, for example — however we contend that it is the discovery and the switching to these different methods (physical technologies) that exhibits the returns.

A well known example of increasing returns arising purely from physical relationships often cited in introductory textbooks¹³ is the geometric relationship between surface area and volume. Suppose a single owner/manager builds an oil pipeline between two points A and B separated by some distance, l . For simplicity, further suppose that the only input is the steel used to construct the pipeline, and the measure of output is the volume of oil that the pipeline can transport over some given period of time. Total input needed is equal to the surface area of the pipeline of radius r , given by $S(r) = 2\pi rl$. The output per period is then the volume of oil that the pipeline can transport given by $V(r) = \pi r^2 l$. By *doubling* the radius from r to $2r$, output can is not merely replicated but *quadrupled* $V(2r) = 4V(r)$, while inputs (the surface area of steel) only double: $S(2r) = 2S(r)$.

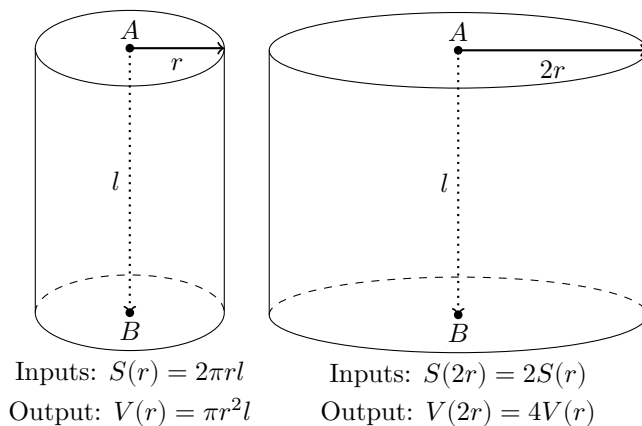


Figure 1: The physical relationship between volume and surface area appears to induce IRS

¹³See e.g. Kaldor (1972), Carlton and Perloff (2005, 38), Church and Ware (2000, 54), Varian (1992, 342).

The pipeline example appears to demonstrate how increasing returns may exist purely due to physical phenomena, in this case the mathematical fact that volume scales faster than surface area.¹⁴ However, upon closer examination, this example actually demonstrates the primacy of institutionally-contingent individual choice, and is not the counterexample to *Proposition 1* it claims to be. This becomes more apparent when we consider why a firm wouldn't merely replicate its production: to increase output, why not simply build *additional* pipelines of radius r ? While it may be possible to achieve a more than proportionate increase in volume by increasing the radius of the pipeline from r to $2r$, it is also possible to simply build *multiple* pipelines with radius r to yield greater V , meaning the physical technology of the pipeline with radius r , and its representative production function, exhibits constant returns.

If physical technologies are truly recipes, then they should be thought of as fully specifying the relevant features of the technology, including in this example the radius of the pipeline (r or $2r$). Therefore, in the strict sense of a recipe, the physical technology that specifies a pipeline with radius r is a *different technology* than one that specifies a pipeline with radius $2r$.

Increasing returns to scale in terms of physical relationships must originate in the deliberate choice of *switching* from one physical technology (recipe) to a technically superior one ($r \rightarrow 2r$). While the greater technical efficiency of a pipeline with a larger radius is indeed the result of a mathematical relationship of volume scaling faster than surface area, the cognitive *realization* that this can be done, and more importantly, the subjective evaluation that it may be more profitable to build a pipeline with a radius $2r$ than to build (e.g.) four pipelines of radius r , is the result of institutionally-contingent choice.

This simple example is not meant to be a definitive proof, but helps clarify our thinking about whether a desired policy outcome depends on policy invariant physical technology or contextual institutional technology to achieve the intended results. The owner-manager's choice is both made possible by, and constrained by, social institutions – such as owning private property and cash flow rights, and relying on the contextual price of steel and the demand for oil to make her decision. In the abstract it would not seem profitable for a firm to produce four pipelines of radius r rather than one pipeline of radius $2r$ to yield $4V(r)$ output, however, in the real world the costs of steel and labor are not the only relevant costs:¹⁵ We could imagine scenarios where, despite it being technologically superior, she would *not* make the decision to double the radius of the pipeline, perhaps due to environmental laws restricting the size of pipelines, or liability rules on oil spills where a larger pipeline would increase expected legal costs, or increased risk of state expropriation,

¹⁴Varian (1992, 342) notes this physical relationship also has its limits “Of course, we can't push this example too far. If we keep doubling the diameter of the pipe, it will eventually collapse of its own weight. Increasing returns to scale usually just applies over some range of output.”

¹⁵We thank one anonymous reviewer for suggesting we make this clarification.

or the political goodwill generated by creating more jobs for building *additional r-sized* pipelines. A change in any of these policies or institutions will impact her production decisions, indicating that IRS is *not* a universal technological feature in this case, but a by-product of various institutions and policies that alter entrepreneurial opportunities.

2.2 Increasing Returns: Institutional Technology

2.2.1 From Profit Maximization to Comparative Institutional Analysis

“The best of all monopoly profits is a quiet life.” - John Hicks (1935, 8)

Any alleged instance of increasing returns to scale must necessarily emerge from the social and institutionally-contingent choices of individuals to use *different physical production technologies*. In some instances, individual entrepreneurs may find it in their interest to alter different physical technologies employing different combinations of inputs, while in others, it may be changes to the algorithm of an existing production technology, that will allow for a more productive use of their existing resources.

This insight compels us to refocus our analytical attention beyond firms and production functions to a broader comparative institutional analysis. Baseline neoclassical models of production assume that the production unit is motivated by profit maximization and is privy to complete information within some given market structure. However, this simplified framework precludes the important effects of particular institutions on the choice of profit-maximizing technology. In an institutionally-barren analysis, the choice of physical technology would make no difference to a firm whether it was a monopolist or it faced a competitive environment, to say nothing of a purely socialist commonwealth. By the behavioral assumption of profit maximization, under conditions of perfect competition, entrepreneurs would find the least-cost efficient methods of production and employ the optimal technological recipe regardless of the market conditions or political environment. This circularly assumes the very behavior that production theory aims to explain: how it is that entrepreneurs come to choose the particular physical technologies that lead to a maximization of allocative efficiency and a minimization of cost (Hayek 1948)! Rather than optimize a fully specified set of technologies under known constraints with complete information, entrepreneurs operate under bounded-rationality and must utilize heuristics to “satisfice” in their decisions (see e.g. Simon (1956), Gigerenzer (2008)). These heuristics can only come from the assistance of the external institutional environment, making for an “ecological rationality” (V. L. Smith 2003).

As such, we can only explain the process of innovation of new technologies by way of rivalrous competition. Even the monopolist who faces the threat of a new entrant that can capture his profits with a new production

technique is compelled to seek his own new technique or lose business. Thus, as Armen A. Alchian (1950) observes, “profit maximization” makes sense less as a behavioral *assumption*, and more as behavior that is observed to *emerge* out of a set of strategies that individual agents employ to successfully adapt to a rivalrous and competitive environment.¹⁶ The competitive environment itself is critical to generating the necessary knowledge of different recipes, as well as incentivizing individuals to produce goods that generate the most economic value. We again cite the literature on “institutional entrepreneurship” (see Footnote 10), which further demonstrates the importance of institutions on entrepreneurial choice: when existing institutions are insufficient to secure property rights or pursue profits, self-interested agents seek to change the institutions in their favor.

It is only with this broader institutional view that Hicks’ quote above makes sense: A monopolist without fear of new entrants would not need to search for new recipes or physical technologies beyond those that merely preserve his returns. Aside from a counterfactual world of greater competition, his existing production method would, by definition, be the most efficient in that market. Thus, he can rest on his laurels knowing that he can resign his business to repeat the same process over and over again, much like the perfect perpetual production process, and retire comfortably. The use of a particular physical technology in production, absent rivalrous competition, could be replicated without end by the push of a metaphorical button, and thus would exhibit constant returns to scale. There is no opportunity for him to exercise entrepreneurial judgment to choose to *switch* to other recipes, which may achieve increasing returns.

In contrast, within a rivalrous environment where multiple producers are competing along various margins, economic survival requires that producers seek (whether intentionally or serendipitously) and ultimately judge it worthwhile to *switch* to more efficient physical technologies. Even an industry with a single firm may allow this outcome in the right institutional setting. Baumol (1982) and Baumol, Panzar, and Willig (1982) demonstrate that “contestable” markets, even with a single incumbent, can act competitively with free entry, shared technology, and no sunk costs. This further highlights the necessary shift of analytical focus to the institutional features of the market firms are operating within.

Producers, each possessing different *internal* institutional technologies, and operating within a shared framework of *external* institutional technologies, will succeed based on whether their institutional technology facilitates that efficient switch between physical technology. This could come as a result of “learning by doing,” where some producers within a firm discover a faster or cheaper algorithm simply from repetition

¹⁶Armen A. Alchian and Kessel (1962) further suggest that a more robust assumption for explaining the behavior of firms is that these agents merely maximize *utility* rather than wealth or profits. This utility can only be maximized by weathering the storm of environmental competition, which from the outside, looks a lot like maximizing profits because only through these attempts can a firm discourage competitors from capturing its market share.

and variation, or from some external “knowledge spillover” that a producer observes from a rival producer and adopts to increase their own productivity. Whether or not a particular production unit can exploit internal learning by doing or external knowledge spillovers will depend on how their institutional technology both facilitates the flow of knowledge from discoverers (e.g. workers on the factory floor, or a rival firm) to those with decision-making power, and the incentive for all parties to discover, share, and implement that knowledge of different physical production technologies.

2.2.2 Institutional Technology at Different Production Scales

Economists use production functions to model production units at different scales: an individual, a firm, an industry, or even a nation. In our description of institutional technology, we wish to highlight the need to disentangle the different institutional technologies present at each level. As mentioned above, whatever level one chooses to model, some institutional technologies exist *internal* to the production unit, and others exist *external* to it. Both aspects are critical: even in the most laissez-faire of economies, some firms are poorly managed; likewise, even in the most restrictive or corrupt polities, some firms thrive. Together, these internal and external constraints determine the information and incentives that shape the choice set that producers face containing the possible physical technologies available for production.

The smallest possible production unit is a single person acting as an owner–manager of resources who uses some physical technology to produce output. Putting aside philosophy of mind problems and assuming there are no *internal* institutional technologies, the individual solely interacts with the external institutions. She must, at minimum, be *alert* to profit opportunities from switching to new technologies (Kirzner 1973) and have the *ability* to implement such a switch. Her entrepreneurial judgment drives the entire production process.

As we ascend in scale from the individual to the firm level, the role of economic organization takes center stage of internal institutional technologies.¹⁷ A large literature on the theory of the firm exists exploring the implications of specializing various aspects of production among persons or organizations, the efficiency of command-and-control hierarchy, and agency costs of operating firms (see e.g. Coase (1937); Stigler (1951); Armen A. Alchian and Demsetz (1972); Fama and Jensen (1983); Williamson (1985)). We emphasize here the internal institutional technologies that enable “firms” to switch between physical technologies (and thus generate increasing returns). As cited in the literature on trade and industrial policy, the main mechanisms are “learning by doing” (Arrow 1962b) investment in research & development (Romer 1990), intellectual property rights (Arrow 1962a), and other spillover effects of knowledge. We however highlight the crucial

¹⁷If we were to continue upwards to the industry, sector, or national level, we would need to add additional organizational and institutional concerns at each stage.

interplay of these valid channels with the above questions of economic organization – how does the firm as a production unit “choose” a particular production technology? Firms must enable those laborers with relevant knowledge gleaned “on the factory floor,” the scientists from its R&D department, or envoys from its trade association partners to bring such knowledge to those with the authority (and incentives) to implement such a choice. Furthermore, once the decision is made by the “top,” it must ultimately be implemented at the “bottom,” though different individuals and specialists within the firm each have their own motivations and limited knowledge. As long as firms have the ability to alter their internal organizational structure, the competitive process identified by Hayek (1948) and Armen A. Alchian (1950) should lead to the survival of those internal institutional technologies that best allow increasing returns to take place. Again, this is not necessarily a conscious process of deliberate optimization by firm managers, but the result would be *as if* it were.

Whatever the level of production chosen by the economist to model, producers exist within specific external institutional technologies that differentially enable them to switch between physical technologies to achieve increasing returns. Economists since A. Smith (1904[1776]) have suggested many key institutions that enable the “great society” under the division of labor which extends the market to further prosperity. David Hume-Hume (2012[1738], bk. III, Part II, §VIII) also famously describes the basic institutional technologies requisite for economic dynamism, “But although men can maintain a small uncultivated society without government, they can’t possibly maintain a society of any kind without justice, i.e. without obeying the three fundamental laws concerning the stability of ownership, its transfer by consent, and the keeping of promises.” Economists have variously highlighted the role of institutions such as clearly defined and enforced property rights (Demsetz 1967; Armen A. Alchian and Demsetz 1973; Acemoglu, Johnson, and Robinson 2001; Acemoglu and Johnson 2005; Acemoglu 2008; Acemoglu and Robinson 2012), rule of law (Frye and Shleifer 1997; Glaeser and Shleifer 2002; La Porta et al. 2004; La Porta, López-de-Silanes, and Shleifer 2008), market prices (Hayek 1945; Kirzner 1973), among other key ingredients. In the broadest terms, in order to maximize the flow and efficient utilization of knowledge, individuals must have access to information and must be able to freely associate in order to optimally organize production and choose physical technologies.

Political authorities at various levels can erect additional constraints on these institutions or substitute something entirely different for them through various policies. Such actions alter the institutional technologies available for producers to be alert to, choose, and implement different physical technologies to increase returns. Poorly understood public policy chosen to “exploit increasing returns” in an industry can conceivably erase their very existence by altering or limiting available institutional technology.

2.3 Summary and Diagrammatic Exposition

For a simple diagrammatic exposition, we recreate and extend Figure 1 in Yoon and Buchanan (1999, 521) below. Let Y represent aggregate output for a composite of inputs X . Production functions for the production of Y using X are represented as lines a_1 and a_2 . Each is a physical technology — a fully specified recipe that — which exhibits constant returns to scale and has constant slope. The dashed portions of a_1 and a_2 are not currently feasible across the range of scales of inputs X . For example, a_1 is not feasible until at least x_1 composite units of X can be utilized. Beyond this threshold, the production functions are solid lines, indicating that output can be increased at constant returns by replicating the production process: doubling the inputs X yields double the output Y .

However, other physical technologies, exemplified by a_2 , are potentially possible. Under the right institutional technology, entrepreneurs can discover and, when threshold x_2 of X is met, switch to physical technology a_2 . This physical technology yields more output Y for a given X than a_1 .

The curve R conceptually shows the relationship of switching between the physical technologies, revealing the presence of increasing returns. R itself is not a production function, nor a set of physical technologies.

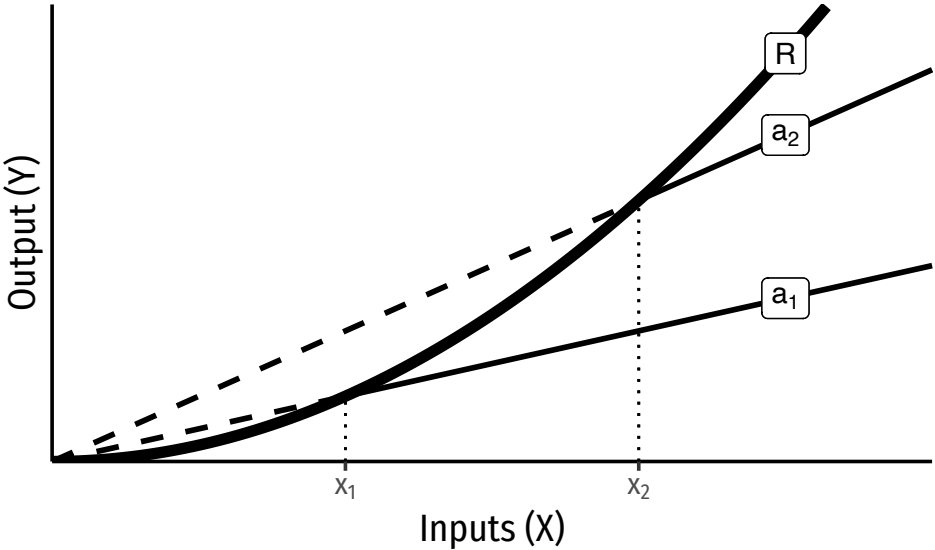


Figure 2: Physical vs. Institutional Technology and Increasing Returns

3 How Policy Affects Physical Technology and Institutional Technology

The trade tariffs described in section 1 proposed by Rodriguez and Rodrik (2001), based on Matsuyama (1992) is just one example of trade policy that aims to spur growth by taking advantage of IRS. Industrial policies such as subsidizing activities or industries that generate IRS have been justified on similar grounds (Rodrik 2004a; Rodrik and Subramanian 2005; Greenwald and Stiglitz 2006; Hausmann, Hwang, and Rodrik 2007). However as we have argued above, for trade tariffs or production subsidies that depend on exploiting IRS to achieve their intended results, we must ensure that they do not significantly alter the institutional technology that provides the incentives and information for the producers to adopt more productive physical technology. We offer two considerations on this matter.

First, a proposed policy will enhance or hinder rivalrous competition. As explained in section 2.2, rivalrous competition is a necessary institutional technology that aligns the incentives of decision makers such that their own pursuits are better accomplished by successively implementing more productive physical technologies. Aside from the effects of tariffs and subsidies on efficiency, these protections decrease the relative cost faced by the productive unit when it fails to learn and adapt. Policies such as trade tariffs, production subsidies, or other regulatory protections from foreign competition, create a less competitive environment that undermines the institutional technology that gives rise to IRS. Increasing returns that are justified through learning by doing are *not* invariant to protectionist policies as these policies undermine the very institutional technology necessary. Productive efficiency will thus diminish to the degree that a policy hinders rivalrous competition among productive units that may feature IRS.

To their credit, most proposals for protectionism or industrial policies based on IRS do consider the static efficiency loss of tariffs and subsidies as a cost of such policies (Rodriguez and Rodrik 2001; Greenwald and Stiglitz 2006). However, they must also consider how these policies interact with the institutional technologies that give rise to productive switching of physical technologies in the first place.

The second aspect that such policies must confront is the dynamic loss of efficiency due to the political economy of such policies. Producers can influence the creation of policies in order to obtain rents by reducing competition (Tullock 1967; Stigler 1971). The availability of rents raises the opportunity cost of discovering and learning new physical technologies since there is an alternative form for firms to raise their private returns (albeit at the expense of the social return) (Baumol 1996; Munger and Villareal-Diaz 2019). These issues are even more pressing in recommending protectionist policies for developing countries since these countries institutions are typically a form of state capitalism (Aligica and Tarko 2012), or crony capitalism

(Aligica and Tarko 2014) where rent-seeking is a dominant concern in economic organization.

The policy creation process does not occur in isolation from producers that can obtain rents. The possibility that such rents could be generated from policy can raise the opportunity cost of switching physical technologies, and thus, damage the institutional technology that would give rise to IRS. Producers may simply seek Hicks' "quiet life" by stifling the competitive process (Kirzner 1985), and divert resources into a "wholly superfluous" process of trying to drive the policy-making process in their favor.

As Tullock (1967) explains, those resources employed by the protected industry in excess of those that would be employed *without* trade protections are a proxy for resources invested in securing protectionist policies. Once we consider (as we should if our model is intended to guide policy) that policy creation is endogenous, these resources invested in lobbying for protection represent a cost not included in any of the present stylized models.¹⁸

Greenwald and Stiglitz (2006, 145) acknowledge that because the benefits of IRS at a *national* scale are rather broad, the "tariffs should be broadly and uniformly applied to industrial products." They also note that such a broad and uniform application would disperse the cost and benefits of the tariffs widely in order to avoid the abuse of such tariffs by special interest groups as described above. However it should be noted that what matters for the logic of concentrated benefits and dispersed costs is the relative, not absolute, size of those that benefit directly from tariffs and those that bear the costs. By definition the group of producers of manufactured goods will be smaller relative to the group of consumers of such goods. Hence, even a broad and uniform application of tariffs over industrial goods will necessarily lead to concentrated benefits and dispersed costs, and the associated logic of collective action (Olson 1971).

It is also worthwhile to question whether a broad and uniform application of protectionist policies is consistent outside of a highly stylized model. Producers of industries that benefit from protection will more easily band together to lobby for protection that goes above and beyond the optimal level by virtue of acting in a less competitive environment. The availability of rents decreases the opportunity cost of not learning new physical technologies. Considering that the opportunity cost of not learning is already smaller in an uncompetitive environment, implementing these protectionist policies in countries where the economic system is already characterized by rent-seeking would only lead to further decrease the opportunity cost of not learning new productive physical technologies. Conversely such protections increase the opportunity cost of not engaging in rent-seeking activities. Justifications for protectionist policies in the form of high theory and highly stylized models provide for cover populist governments to perpetuate the rent-seeking status quo in countries that

¹⁸For a thorough theoretical treatment that conceptualizes both productive economic activities and the creation of policies that regulate them as two different tools that individuals utilize to pursue their objectives see Wagner (2014), and A. Smith, Wagner, and Yandle (2011) for applications of this theoretical framework.

desperately need to improve their institutions in order to escape poverty.

3.1 Case Studies of “Successful” Infant Industries in the 19th Century United States

Below, we provide two examples of apparent success stories of infant industries in 19th century United States — that is, cases of an industry that initially had a comparative disadvantage, but was somehow “destined” to gain comparative advantage (through learning-by-doing, positive spillovers, etc), that the State provided temporary tariff protection that enabled the industry to reach the scale necessary to attain the advantage (and then the tariffs were repealed): Irwin (2000) which examines late-19th century U.S. domestic tinplate manufacturing and Head (1994) which examines steel rail production in the same era.

Unlike most manufacturing industries in the postbellum U.S., tinplate manufacturing stood on its own without protection until the McKinley tariff 1890 (Irwin 2000, 336). Following the tariff, which encouraged entry of domestic firms, industry output greatly increased despite tariff reductions, and price ultimately fell so low that the U.S. became a net exporter (ibid, 336). Irwin (2000, 336) suggests that to properly understand this apparent case of a successful infant industry tariff must consider the counterfactual of the industry’s development without such protection, (335). The analysis by Taussig (1915, 151) suggests “the same sort of growth would doubtless have taken place eventually, tariff or no tariff; but not so soon or on so great a scale.” Irwin (2000, 336–37) suggests “tinplate was not an ‘infant industry’ that floundered due to the lack of previous production experience, but rather one in which domestic production was not profitable until the high domestic cost of iron and steel inputs fell to levels that roughly matched those of foreign rivals. In the absence of the McKinley tariff, the U.S. tinplate industry would have established itself about a decade later as those material input costs converged with those abroad.” In other words, the need was for the industry to reach economies of scale to price competitively relative to foreign imports.

The industry was ultimately able to attain economies of scale largely due to “both domestic and international technological and learning-based knowledge spillovers” (Irwin 2000, 335). Irwin quotes a pro-tariff argument by the president of the American Tin Plate Association as claiming “after we get 50 mills in this country and exchange our ideas we can reduce the price by the use of improved machinery and methods which they never thought of in the other countries.” Much of the positive spillovers came from Welsh immigrants who were highly skilled tinplate workers that brought embodied knowledge of Welsh production methods. This allowed the U.S. industry to rapidly catch up to the production possibilities frontier, but afterwards, domestic learning-by-doing allowed the U.S. to expand the frontier with domestically discovered techniques. Irwin (2000, 346) quotes Berthoff (1953, 68–69): “Within a few years, however, as equipment and methods were

improved, Welsh workmen no longer were necessary nor, over-fond of time-honored ways, even desirable.” Notably, Irwin (2000, 346) states that “That U.S. tinsplate entrants could draw upon Welsh production techniques by hiring the skill and expertise of Welsh tanners suggests that they did not face insurmountable hurdles to entering the industry.”

Head (1994) examines another apparent success case of tariffs protecting the infant U.S. steel rail industry in the late 19th century. The U.S. was initially at a comparative disadvantage, with costs high relative to England, despite it being one of the largest post-Civil War industries (143). Following tariffs, this position was ultimately reversed by 1913, with the U.S. becoming a key exporter (and negligible importer) through the 20th century, even well after the tariffs were dropped (ibid, 143). However, Head (1994, 143) notes “discoveries of new natural resources (iron ore and coal) and the construction of infrastructure (railroads and canals) probably shifted the U.S. comparative advantage towards metal production.” His simulations suggest that if “the United States had not had the ore discoveries and improvements in infrastructure that caused its pig iron prices to decline relative to British pig iron prices, [would] protection-induced learning have been enough to dominate the standard determinants of comparative advantage? The answer seems to be no,” (160). His simulations also emphasize the role of learning-by-doing, without which, “[a]fter a few years of small production levels, the domestic industry virtually disappears, crushed by falling British prices,” (161). Thus, “Steel rails appears to be an industry where both factor endowments and dynamic returns to scale played vital roles. Under both constant relative pig iron prices and no learning the steel rail tariff causes substantial welfare losses: 1,108 and 194 million 1992 dollars,” (161).

However, like the case of tinsplate manufacturing, much of the learning-by-doing success of domestic steel rail industry was due to immigration of skilled workers. The model used by Head (1994) assumes that learning-by-doing fully spill over between domestic firms but not across countries. Irwin (2000, Fn 27) notes this assumption would ensure that entrants would never catch up to their British rivals. However, the period was “one of tremendous international labor mobility and the United States was the principal destination for skilled European migrants” (ibid, Fn 27). British iron officials visiting the U.S. reported in 1901 that “a considerably number of the heads of the American iron industry of today acquired their training, their knowledge, and their experience in British works,” (Berthoff 1953, 67).

These examples of apparently successful use of trade and industrial policy to exploit IRS can be better clarified using the framework we have outlined above. The reason the policies “worked” was largely due to the institutions in place that enabled entrepreneurs to discover new physical technologies — such as the Bessemer process in steel-making [Head1994,143] — and find it profitable to switch to them, rather than replicating their own existing technologies.

One of the primary reasons the learning-by-doing was able to take place and the infant industries able to reach sufficient scale was due to liberal immigration laws, and other institutions of economic freedom that made the growing United States attractive to foreign workers and capital investment during the period. We could imagine one of many counterfactual institutional arrangements where e.g. U.S. immigration laws were more stringent, and without the importation of Welsh techniques and skills in tinsplate manufacturing, or British steel rail producing, these industries would not have been as successful in attaining larger scale and lower costs.

Bessen (2015) suggests that most technological change is embodied in skilled workers, and requires institutions such as trade associations, industry standards, and a robust labor market for technological change to stimulate productivity growth and higher wages. When standards become common in an industry and robust labor markets emerge, competition between firms in the industry increases the expected return to workers investing in non-firm-specific human capital, and raises their opportunity cost of working at a low-productivity firm (as they can work at a more efficient competitor for higher wages). On average, this process from the initial discovery of various tinkerers to commercially-viable new products and mature industries takes about 30 years (ibid, 39-40).

Many industries may *retroactively* be shown to exhibit increasing returns. It is much more difficult to identify such industries in advance, and stake public (or private, for that matter) funds on them. In the examples of tinsplate and steel rail manufacturing, both Irwin (2000) and Head (1994) are cautious to suggest that the infant industries argument is lent little credibility by these historical cases. In both cases, confounding variables (such as immigration, discovery of other natural resources, and growth of the American economy) contributed to the success of these industries, in spite of, or despite, tariff protection. Ultimately, Irwin's simulations ultimately suggests that "the McKinley duty fails to pass a cost-benefit test," (336). Head (1994, 161) notes that "under such seemingly favorable conditions, the protectionist policy failed to yield more impressive welfare improvements." He considers the possibilities that Congress extended protection for too long, or that they used the wrong policy for protection, noting the former proved more damaging to welfare gains than the latter (161—162). Such findings, even in the most favorable of cases, suggest extreme caution for identifying *ex ante* industries that should be protected to exploit increasing returns.

4 Implications

Increasing returns can be understood as an implication from Smith's (1776, Chs. I-III) observations about the effects of the division of labor on productivity and its limits due to the extent of the market. Many have

made light of these famous observations, emphasizing the gains to specialization *within* a set of institutions or production process. Smith’s pin factory serves as the jumping-off point for a theory of how specialization within a particular production process can increase output. However, since Young (1928); many have pointed out that the main insights for economic growth that come from Smith’s discussion of the division of labor are at a much more *generalized* level than a single pin factory (Stigler 1951; Buchanan 1994; Yoon and Buchanan 1999; Buchanan and Yoon 2000; @ Boettke and Candela 2017): it is the importance of particular institutions that enable the extent of the market to increase, making sustained economic growth feasible.¹⁹ Entrepreneurs are able to discover new technologies and specialize more at these larger scales.

We have argued that the application of the theory of increasing returns to policy suffers from a fundamental misunderstanding about the nature of the production function and actual origin of increasing returns: They are not fundamental features of physical production, but are emergent features of individual choice within a specific institutional technology that provides the incentives for *discovering* and *switching* between physical technologies to take place. As such, IRS still serve to create broad generalizations to explain economic growth, but we must take special care when they are invoked to promote specific policies.

The clarification between the role of physical technology and institutions in increasing returns provides insights into many fields and active debates in economics beyond trade and industrial policy.²⁰ It has implications for the burgeoning literature of public economics and responses to market failures²¹. Consider any public good that has positive spillovers — whether art or an invention, public services, etc. The marginal social benefit of the good being produced is significant, and larger than the private marginal benefits to producers. If the private marginal cost faced by producers is sufficiently high, there is no incentive for private producers to supply the good, to the detriment of social welfare more broadly. Whatever goods get produced, if any at all, are produced under constant returns to scale, generating no spillovers. However, if entrepreneurs can find a way to supply the good profitably, including altering the rules or institutions, to discover a technology that will raise the private benefit beyond the private cost, they will supply the good despite positive spillovers.²² Hence, institutions that allow discovery and changing of physical technology enable increasing returns and positive spillovers from providing the public goods.²³

¹⁹Yoon and Buchanan (1999) come closest to briefly making our point: “If there are advantages of specialization, however, an expansion in scale of the economy as a whole will not take place within the parametric constraint of a given technology. The extended specialization dictates that efficient operation requires a shift to a new technology, to a *different production function*,” [518, emphasis in original].

²⁰We thank one anonymous reviewer for emphasizing our implications in these literatures and for the alternate framing of our approach described here.

²¹See e.g. Cowen (2018[1988]) for a collection of case studies

²²Examples abound in the literatures on the private supply of e.g. beekeeping services (Cheung 1973), law and order (Friedman 1979; Benson 1989; Ellickson 1994), or lighthouses (Coase 1974; Candela and Geloso 2018, 2019)

²³This insight further has applications to the literature on state capacity and its role in economic development (see e.g. Dinuccio 2017; Johnson and Koyama 2017; Geloso and Salter 2020). If the provider of these goods is a State actor rather than a private producer, it is in the interest of the ruler(s) to find the institutions that enable such discovery, which will unleash

This article has provided a theoretical framework that should be readily applied to clarify historical and present cases of suspected increasing returns industries and policies designed to create or exploit them. We hope that future work will attempt to estimate the marginal contributions of various institutions on the sources of learning in production that features IRS. We await further empirical testing that can separate and verify the primary effect that alternative institutional forms have on increasing returns.

5 Acknowledgments

The authors would like to thank Peter Lewin, Diana Thomas, Austin Middleton, Virgil Storr, the participants of the Southern Economic Association Annual Conference, and two anonymous reviewers for their helpful comments on earlier drafts. All remaining errors are our own.

6 References

- Acemoglu, Daron. 2008. "Oligarchic Versus Democratic Societies." *Journal of the European Economic Association* 6 (1): 1–44.
- Acemoglu, Daron, and Simon Johnson. 2005. "Unbundling Institutions." *Journal of Political Economy* 113: 949–94.
- Acemoglu, Daron, Simon Johnson, and James Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91: 1369–1401.
- . 2002. "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution." *Quarterly Journal of Economics* 117: 1231–94.
- Acemoglu, Daron, and James A. Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. New York: Crown Business.
- Alchian, Armen A. 1950. "Uncertainty, Evolution, and Economic Theory." *Journal of Political Economy* 58 (3): 211–21.
- Alchian, Armen A., and Harold Demsetz. 1973. "The Property Rights Paradigm." *Journal of Economic History* 33 (1): 16–27.
- Alchian, Armen A., and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *American Economic Review* 62 (5): 777–95.
- Alchian, Armen A., and Reuben A. Kessel. 1962. "Competition, Monopoly, and the Pursuit of Pecuniary Gain." In *Aspects of Labor Economics*, 157–84. Princeton University Press.
- Aligica, Paul Dragos, and Vlad Tarko. 2012. "State Capitalism and the Rent-Seeking Conjecture." *Constitutional Political Economy* 23 (4): 357–79.
- . 2014. "Crony Capitalism: Rent Seeking, Institutions and Ideology." *Kyklos* 67 (2): 156–76.
- Arrow, Kenneth J. 1962a. "Economic Welfare and the Allocation of Resources for Invention." In *The Rate and Direction of Inventive Activity: Economic and Social Factors*, edited by UMI. Universities-National Bureau.
- . 1962b. "The Economic Implications of Learning by Doing." *The Review of Economic Studies* 29 (3): pp. 155–173. <http://www.jstor.org/stable/2295952>.
- Arthur, W. Brian. 1989. "Competing Technologies, Increasing Returns, and Lock-in by Historical Events." *The Economic Journal* 99: 116–31.
- . 1994. *Increasing Returns and Path Dependence in the Economy*. Ann Arbor, MI: University of Michigan Press.

increasing returns within the society and create more taxable income for the rulers.

- . 1996. “Increasing Returns and the New World of Business.” *Harvard Business Review*.
- Babbage, Charles. 1835. *On the Economy of Machinery and Manufactures*. 4th ed. London: Charles Knight.
- Barro, Robert J., and Xavier Sala-i-Martin. 2004. *Economic Growth*. Second Edition. MIT Press.
- Baumol, William J. 1982. *Contestable Markets: An Uprising in the Theory of Industry Structure*. *American Economic Review*. Vol. 72. 1.
- . 1996. *Entrepreneurship: Productive, Unproductive, and Destructive*. *Journal of Business Venturing*. Vol. 11.
- Baumol, William J., John C. Panzar, and Robert D. Willig. 1982. *Contestable Markets and the Theory of Industry Structure*. New York: Harcourt Brace Jovanovich, Inc.
- Benson, Bruce L. 1989. “The Spontaneous Evolution of Commercial Law.” *Southern Economic Journal* 5: 644–61.
- Berthoff, Rowland T. 1953. *British Immigrants to Industrial America, 1750-1950*. Cambridge, MA: Harvard University Press.
- Bessen, James. 2015. *Learning by Doing: The Real Connection Between Innovation, Wages, and Wealth*. New Haven, CT: Yale University Press.
- Boettke, Peter J, and Rosolino A Candela. 2017. “The Liberty of Progress: Increasing Returns, Institutions, and Entrepreneurship.” *Social Philosophy and Policy* 34 (2): 136–63.
- Buchanan, James M. 1994. “The Return to Increasing Returns: An Introductory Summary.” In *The Return to Increasing Returns*, edited by James M Buchanan and Yong J Yoon. University of Michigan Press.
- Buchanan, James M, and Yong J Yoon. 2000. “A Smithean Perspective on Increasing Returns.” *Journal of the History of Economic Thought* 22 (01): 43–48.
- Candela, R. A., and V. J. Geloso. 2018. “The Lightship in Economics.” *Public Choice* 176 (3-4): 479–506.
- . 2019. “Why Consider the Lighthouse a Publicgood?” *International Review of Law and Economics* 60: 105852.
- Cardinale, Ivano. 2018. “Beyond Constraining and Enabling: Toward Microfoundations for Institutional Theory.” *Academy of Management Review* 43 (1): 132–55.
- Carlton, Dennis W., and Jeffrey M. Perloff. 2005. *Modern Industrial Organization*. 4th ed. New York: Addison-Wesley.
- Cheung, Steven NS. 1973. “The Fable of the Bees: An Economic Investigation.” *Journal of Law and Economics* 16: 11–33.
- Church, Jeffrey, and Roger Ware. 2000. *Industrial Organization: A Strategic Approach*. New York: Irwin McGraw-Hill.
- Clark, John Bates. 1889. “Possibility of a Scientific Law of Wages.” *Publications of the American Economic Association* 4 (1): 39–69.
- . 1891. “Distribution as Determined by a Law of Rent.” *Quarterly Journal of Economics* 5 (3): 289–318.
- Coase, Ronald H. 1937. “The Theory of the Firm.” *Economica* 4: 386–405.
- . 1974. “The Lighthouse in Economics.” *Journal of Law and Economics* 17 (2): 357–76.
- Cowen, Tyler, ed. 2018[1988]. *Public Goods and Market Failures: A Critical Examination*. New York: Routledge.
- Demsetz, Harold. 1967. “Toward a Theory of Property Rights.” *The American Economic Review* 57 (2): 347–59.
- Dincecco, Mark. 2017. *State Capacity and Economic Development*. Cambridge, MA: Cambridge University Press.
- Ellickson, Robert. 1994. *Order Without Law: How Neighbors Settle Disputes*. Cambridge, MA: Harvard University Press.
- Fama, Eugene F, and Michael C Jensen. 1983. “Separation of Ownership and Control.” *Journal of Law and Economics* 26 (2): 301–25.
- Friedman, David D. 1979. “Private Creation and Enforcement of Law: A Historical Case.” *Journal of Legal Studies* 8: 399–415.
- Frye, Timothy, and Andrei Shleifer. 1997. “The Invisible Hand and the Grabbing Hand.” *American Economic Review* 87 (2): 354–58.
- Geloso, Vincent J, and Alexander W Salter. 2020. “State Capacity and Economic Development: Causal

- Mechanism or Correlative Filter?" *Journal of Economic Behavior and Organization* 170: 372–85.
- Gigerenzer, Gerd. 2008. *Rationality for Mortals*. Oxford, UK: Oxford University Press.
- Glaeser, Edward L., Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2004. "Do Institutions Cause Growth?" *Journal of Economic Growth* 9: 271–303.
- Glaeser, Edward L., and Andrei Shleifer. 2002. "Legal Origins." *Quarterly Journal of Economics* 117 (4): 1193–1229.
- Graham, Frank D. 1923. "Some Aspects of Protection Further Considered." *The Quarterly Journal of Economics*, 199–227.
- Greenwald, Bruce, and Joseph E Stiglitz. 2006. "Helping Infant Economies Grow: Foundations of Trade Policies for Developing Countries." *The American Economic Review* 96 (2): 141–46.
- Hausmann, Ricardo, Jason Hwang, and Dani Rodrik. 2007. "What You Export Matters." *Journal of Economic Growth* 12 (1): 1–25.
- Hayek, Friedrich A. 1945. "The Use of Knowledge in Society." *The American Economic Review* 35 (4): 519–30.
- . 1948. "The Meaning of Competition." *Individualism and Economic Order*, 92–106.
- Head, Keith. 1994. "Infant Industry Protection in the Steel Rail Industry." *Journal of International Economics* 37: 141–65.
- Hicks, John R. 1935. "Annual Survey of Economic Theory: The Theory of Monopoly." *Econometrica* 3 (1): 1–20.
- Hume, David. 2012[1738]. *A Treatise of Human Nature*. Project Gutenberg.
- Irwin, Douglas A. 2000. "Did Late-Nineteenth-Century u.s. Tariffs Promote Infant Industries? Evidence from the Tinplate Industry." *Journal of Economic History* 60 (2): 335–60.
- Johnson, Noel D, and Mark Koyama. 2017. "States and Economic Growth: Capacity and Constraints." *Explorations in Economic History* 64: 1–20.
- Kaldor, Nicholas. 1972. "The Irrelevance of Equilibrium Economics." *The Economic Journal* 82 (328): 1237–55.
- Kirzner, Israel M. 1973. *Competition & Entrepreneurship*. Chicago: University of Chicago Press.
- . 1985. *Discovery and the Capitalist Process*. Chicago, IL: Chicago University Press.
- Knight, Frank H. 1924. "Some Fallacies in the Interpretation of Social Cost." *The Quarterly Journal of Economics* 38 (4): 582–606.
- . 1925. "On Decreasing Cost and Comparative Cost: A Rejoinder." *The Quarterly Journal of Economics* 39 (2): 331–33.
- Krugman, Paul. 1979. "Increasing Returns, Monopolistic Competition, and International Trade." *Journal of International Economics* 9: 469–79.
- . 2009. "The Increasing Returns Revolution in Trade and Geography." *American Economic Review* 99 (3): 561–71.
- La Porta, Rafael, Florencio López-de-Silanes, Cristian Pop-Eleches, and Andrei Shleifer. 2004. "Judicial Checks and Balances." *Journal of Political Economy* 112 (2): 445–70.
- La Porta, Rafael, Florencio López-de-Silanes, and Andrei Shleifer. 2008. "The Economic Consequences of Legal Origins." *Journal of Economic Literature* 46 (2): 285–332.
- Leeson, Peter T., and Peter J. Boettke. 2009. "Two-Tiered Entrepreneurship and Economic Development." *International Review of Law and Economics* 29: 252–59.
- Lucas, Robert E. 1976. "Econometric Policy Evaluation: A Critique." In *Carnegie-Rochester Conference Series on Public Policy*, 1:19–46. Elsevier.
- Machlup, Fritz. 1937. "On the Meaning of Marginal Product." In *Explorations in Economics: Notes and Essays Contributed in Honor of f. W. Taussig*, edited by Frank Taussig. New York: McGraw-Hill.
- Marshall, Alfred. 1920. *Principles of Economics: An Introductory Volume*. Philadelphia, PA: Porcupine.
- Martin, Adam, and Diana W. Thomas. 2013. "Two-Tiered Political Entrepreneurship and the Congressional Committee System." *Public Choice* 154: 21–37.
- Matsuyama, Kiminori. 1992. "Agricultural Productivity, Comparative Advantage, and Economic Growth." *Journal of Economic Theory* 58 (2): 317–34.
- Munger, Michael C, and Mario Villareal-Diaz. 2019. "The Road to Crony Capitalism." *The Independent Review* 23 (3): 331–44.
- North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cam-

- bridge University Press.
- Olson, Mancur. 1971. *The Logic of Collective Action: Public Goods and the Theory of Groups, Second Printing with New Preface and Appendix (Harvard Economic Studies)*. Harvard University Press.
- Rodriguez, Francisco, and Dani Rodrik. 2001. "Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence." In *NBER Macroeconomics Annual 2000, Volume 15*, 261–338. MIT Press.
- Rodrik, Dani. 2004a. "Industrial Policy for the Twenty-First Century." KSG Working Paper No. RWP04-047.
- . 2004b. "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development." *Journal of Economic Growth* 9: 131–65.
- Rodrik, Dani, and Arvind Subramanian. 2005. "From "Hindu Growth" to Productivity Surge: The Mystery of the Indian Growth Transition." *IMF Staff Papers* 52 (2): 193–228.
- Romer, Paul M. 1986. "Increasing Returns and Long-Run Growth." *The Journal of Political Economy* 94 (5): 1002–37.
- . 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5): S71–102.
- Safner, Ryan. 2016. "Institutional Entrepreneurship, Wikipedia, and the Opportunity of the Commons." *Journal of Institutional Economics*, no. Forthcoming.
- Scazzieri, Roberto. 2014. "A Structural Theory of Increasing Returns." *Structural Change and Economic Dynamics* 29: 75–88.
- Simon, Herbert A. 1956. "Rational Choice and the Structure of the Environment." *Psychological Review* 63 2: 129–38.
- Smith, Adam. 1904[1776]. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Edited by Edwin Cannan. 5th ed. London: Methuen & Co., Ltd.
- Smith, Adam, Richard E Wagner, and Bruce Yandle. 2011. "A Theory of Entangled Political Economy, with Application to TARP and NRA." *Public Choice* 148 (1-2): 45–66.
- Smith, Vernon L. 2003. "Constructivist and Ecological Rationality in Economics." *American Economic Review* 93 (3): 465–508.
- Stigler, George J. 1951. "The Division of Labor Is Limited by the Extent of the Market." *Journal of Political Economy* 59 (3): 185–93.
- . 1971. "The Theory of Economic Regulation." *The Bell Journal of Economics and Management Science* 2 (1): 3–21.
- Stiglitz, Joseph E. 2014. "Creating a Learning Society." *Project Syndicate*.
- Stiglitz, Joseph E, and Bruce C Greenwald. 2014. *Creating a Learning Society: A New Approach to Growth, Development, and Social Progress*. New York: Columbia University Press.
- Taussig, Frank W. 1915. *Some Aspects of the Tariff Question*. Cambridge, MA: Harvard University Press.
- Tullock, Gordon. 1967. "The Welfare Costs of Tariffs, Monopolies, and Theft." *Economic Inquiry* 5 (3): 224–32.
- Varian, Hal R. 1992. *Microeconomic Analysis*. Third. W. W. Norton & Company, Inc.
- Wagner, Richard E. 2014. "Entangled Political Economy: A Keynote Address." *Entangled Political Economy (Advances in Austrian Economics, Volume 18) Emerald Group Publishing Limited* 18: 15–36.
- Wicksteed, Phillip H. 1894. *Essay on the Co-Ordination of the Laws of Distribution*. London: L.S. E. Reprint (1932).
- Williamson, Oliver E. 1985. "The Economic Institutions of Capitalism."
- Yoon, Yong J, and James M Buchanan. 1999. "Generalized Increasing Returns, Euler's Theorem, and Competitive Equilibrium." *History of Political Economy* 31 (3): 511–23.
- Young, Allyn. 1928. "Increasing Returns and Economic Progress." *The Economic Journal* 38 (152): 527–42.