Understanding Growth through Automation*

Lukasz A. Drozd
Federal Reserve Bank of Philadelphia

Mathieu Taschereau-Dumouchel
Cornell University

Marina M. Tavares
International Monetary Fund

September 19, 2021

Abstract

Uzawa's theorem and the Kaldor facts tell us that the aggregate production function must be Cobb-Douglas when there is capital-augmenting technical change. On the micro level we observe the phenomenon of *growth through automation*—growth brought about by capital-augmenting technical change that results in a steady displacement of labor by capital across tasks or activities. The neoclassical growth theory tells us there is no contradiction between these phenomena, but the approach of using a reduced form (Cobb-Douglas) aggregate production function obscures the connection between them and does not tell us what automation is or does. Here we enrich the neoclassical theory by developing a tractable task-based theory of automation that links growth, automation, and factor share dynamics. We use it to derive general conditions under which *growth through automation* is consistent with balanced growth and back out the underlying task technology. We characterize advancements in automation technology that are labor share displacing when factor shares are price-invariant.

Key words: Automation, labor share, Uzawa's theorem, Cobb-Douglas production function, capital-augmenting technological progress, balanced growth

JEL Classifications: D33, E25, O33, J23, J24, E24, O4.

*The views expressed herein are solely those of the authors and do not necessarily reflect those of the International Monetary Fund, the Federal Reserve Bank of Philadelphia or the Federal Reserve System. Drozd (corresponding author): lukasz.drozd@phil.frb.org, Research Department, Federal Reserve Bank of Philadelphia Ten Independence Mall Philadelphia, PA 19106-1574; Taschereau-Dumouchel: Department of Economics, Cornell University, mt763@cornell.edu. Mendes Tavares: Research Department, International Monetary Fund, MMendesTavares@imf.org. The paper has been inspired by Guido Menzio's Society of Economic Dynamics 2021 plenary address. All errors are ours.

One of the fundamental challenges in macroeconomics is to describe the aggregate production function in a way that is both tractable and realistic. The neoclassical theory's success in becoming the foundation of modern macroeconomics largely stems from the fact that it effectively addresses this problem. By the Kaldor facts (Kaldor, 1961; Jones and Romer, 2010), macroeconomic models are required to be consistent with the existence of a balanced growth path. By the Uzawa's steady-state growth theorem (Uzawa, 1961), and the presence of capital-augmenting technological progress contributing to growth (Greenwood et al., 1997; Grossman et al., 2017), the aggregate production function must be Cobb-Douglas. In that case, a single observable statistic pins down the production function of the entire economy: the labor's share in the national income. As Jones and Romer (2010) once put it, "there is no longer any interesting debate about the features that a model must contain to explain. These features are embedded in one of the great successes of growth theory in the 1950s and 1960s, the neoclassical growth model."

This result has been fundamental to macroeconomics for decades, but the recent evidence increasingly tests economists' confidence in its central tenet: the stability of the labor share. For example, Karabarbounis and Neiman (2013) measure the decline in multiple countries and conclude that the labor share has been falling globally since the 1980s.[1] Several other studies provide evidence that the decline is spreading to a growing number of sectors.[2] What further reinforces this concern is the fact that labor share declines have been also associated with a changing nature of technology brought about by advances in information technology (IT) and dexterous automation. According to this hypothesis, the declining price of ever more capable machines augmented by IT results in *growth though automation* that is now eating into the labor share.[3] These developments have spurred renewed interest in the microfoundations of the aggregate production function to better understand the connections between growth, automation, and factor share stability (Acemoglu and Restrepo, 2018). The goal of the current paper is to reexamine this issue from the perspective of a neoclassical theory augmented by a task-based theory of production that can speak to the phenomenon of automation

[1]See Dao et al. (2017) for updated evidence. See also Gomme and Rupert (2004) and Armenter (2015) for an accessible review of the measurement of the labor share in the US. Koh et al. (2020) and Gutierrez and Piton (2020) discuss importantant measurement issues associated with the labor share, which is a competing hypothesis.

[2]See sectoral evidence in Elsby et al. (2013) and Manyika et al. (2019).

[3]Popular writings by Brynjolfsson and McAfee (2014), Ford (2009) and Frey (2020) provide an accessible overview how modern technology is reshaping production. A number of academic papers allude to the transformative effect of modern automation technology. Some relate it to the labor share and almost all to labor demand. See, for example, the work by Graetz and Michaels (2018), Acemoglu and Restrepo (2018), Acemoglu and Restrepo (2020), Autor and Salomons (2018), Manyika et al. (2019), Muro et al. (2019), Arntz et al. (2016), Autor et al. (2003), Goos and Manning (2007), Acemoglu and Autor (2011), David and Dorn (2013), Michaels et al. (2014) and Dvorkin and Monge-Naranjo (2019).

in a descriptively realistic manner.

The specific question we ask is whether *growth through automation*—which we define as growth brought about by technical progress that results in a steady displacement of labor by capital—is consistent with the existence of a balanced-growth path in such a framework. Key to this inquiry is Uzawa's steady-state growth theorem, which asserts that factor shares must be factor price-invariant and the aggregate production function must be Cobb-Douglas. The neoclassical theory tells us that this is possible, but the answer comes out of a reduced form (Cobb-Douglas) production function that obscures the connection between these phenomena and thus does not provide a descriptively realistic explanation for the coexistence of automation and Kaldor facts. This prevents us from understanding how advancements to automation technology might be reshaping production, and whether this could lower the labor share when Uzawa's theorem applies. The concerns about automation eating into the labor share derive precisely from this failure of the neoclassical theory.

To address this shortcoming, here we develop a novel task-based theory of production and use it to examine the general conditions on the specification of the task technology that can yield Cobb-Douglas production function in aggregate. The starting point of our analysis is the observation that Uzawa's theorem does not imply that factor shares are perfectly constant amid imperfect measurement, nor that they are exogenous, only that they must be sufficiently stable to explain past data. The goal is to focus on such a condition to back out the underlying task technology and this way learn what the future might hold for labor.

Our task model is inspired by prior work by Acemoglu and Restrepo (2018) (AR18 hereafter), but it is different from theirs in that it does not rely on the ever expanding task space to ensure balanced growth. Unlike in AR18, the task space is large but static, and growth is at least in part generated by capital augmenting technological progress of some sort (we consider several examples). Production requires completing a random subset of tasks varying in complexity on the real line, which is determined once and for all upon the inception of a plant—the basic unit of production in the model. Tasks differ in the productivity of capital depending on complexity, which determines the use of capital versus labor. Complexity corresponds to the measure of tasks that must be completed to produce a piece of capital specific to a particular task (complexity). As in AR18, labor and capital are perfect substitutes at the task level, resulting in a cutoff complexity below which all tasks are completed using capital. As the price of capital declines, this cutoff increases and an ever increasing set of tasks gets automated, which we earlier called growth through automation.

Our main result shows that the Cobb-Douglas aggregate production function obtains *if and only* *if* task complexity follows a power law on the real line and capital productivity relative to labor across tasks is of constant elasticity with respect to task complexity. A simple implementation of this result is a Pareto distribution with a positive lower bound on complexity, as is required by such a distribution. As we show, this achieves the result approximately and the approximation is arbitrarily fine and asymptotically exact as the economy grows. The exact result requires that the lower bound of the distribution extends down to zero, which is a limit result and converges to a well-defined limit economy endowed with an infinite measure based off the Pareto probability distribution (a finite measure). The limit economy delivers balanced growth exactly because it aggregates exactly to the Cobb-Douglas production function. The mechanism that keeps the labor share from falling in the course of growth through automation is the degree of diminishing returns from automation across tasks; that is, how steeply the task complexity is rising in distribution. This result does not require the limit economy. As we show, any globally steeper schedule than the one associated with Cobb-Douglas production function under the standard Pareto distribution will do.[4] The infinite measure is what makes capital essential and ensures the Inada condition, no more. This leads us to the conclusion that the dynamics of factor share in the course of growth through automation and the balanced growth are separate phenomena according to our theory, despite the latter demanding a Cobb-Douglas production function. Put differently, Uzawa's result is more about the global properties of Cobb-Douglas production that are required for balanced growth than its local properties that prevent the labor share from falling at a point in time.

As for the labor share declines associated with advancements in automation technology, our model points to innovations that make capital better in dealing with task complexity globally on the task domain (imply a flatter capital requirement). Such innovations lead to a permanently lower labor share and yet can preserve the Cobb-Douglas aggregate production function. We identify such a class of technologies analytically and discuss their key characteristics in light of the kind of transformative technologies that are believed to have taken automation to the next level, such as artificial intelligence (AI) or information and communication technology (ICT). The main effect of inventing one of such technologies, which we refer to as automation breakthrough, is that at a small fixed cost of application it "compresses" the task space in the production of capital goods for the more complex tasks. As we explained, Cobb-Douglas aggregate production function obtains in our model when the elasticity of capital requirement per task with respect to task complexity is constant and the distribution is Pareto, with the coefficient of proportionality corresponding to the exponent

[4] Steeper here means globally higher second derivative.

of the implied Cobb-Douglas production function. The diffusion of automation breakthrough across the economy endogenously changes that coefficient.

Let us conclude by invoking the AR18's line of inquiry inspired by the Leontief's analogy between horses and humans. They asked: What differentiates humans from horses so that the former will not share the fate of the latter of being displaced by machines? Our answer is different: unlike horses, humans are fungible across a vast array of existing and widely heterogeneous tasks, and that alone can be enough of a defense line for labor to maintain its share in income. This, however, does not mean that humans are invincible to technical progress. As we mentioned, breakthrough technologies may reduce the labor share, and such technologies exhibit features that are worryingly similar to the kind of innovations that are considered the catalysis of the current wave of automation, such as artificial intelligence (AI) or information and communication technology (ICT).[5] While the answer whether these technological advancements correspond to a breakthrough is beyond the scope of our paper, Section 5 discusses several examples that are suggestive of an affirmative answer, opening the floor for a debate guided by theory.

Related literature. As mentioned, our work is closely related to AR18. In particular, we build here on the key premise of their analysis, which we earlier referred to as growth through automation. But our approach and our model are fundamentally different. In particular, we do not assume a technology that results in an expansion of the task space to ensure a balance growth path and explore a flexible but conceptually simpler static setup. As a result, while AR18 show that on the balanced growth path factor shares are constant despite growth through automation, their aggregation result does not (generically) imply that the production function is Cobb-Douglas, nor that factor shares are factor price-invariant.[6] Our model does not include such a feature and asks what we can learn by exploring a more stringent route of aggregation to a Cobb-Douglas production function. We also entertain a relaxed set of assumptions to yield an approximate balanced-growth path.[7] The main insight from our analysis in the context of their work is that, to obtain approximate balanced growth, or the property that the labor share does not decline in the course of automation, such a feature is not needed. On the other hand, the exact balanced growth path result vindicates their approach showing that such or related feature is actually indispensable. However, it can modeled as a static

[5] See, for example, the popular writing by Brynjolfsson and McAfee (2014), Ford (2009) and Frey (2020).

[6] Such a result in their model requires Cobb-Douglas aggregation of tasks.

[7] On the balance growth path the two theories are distinguished by the extent to which factor shares are factor price invariant in short- and medium run. Price invariant factor shares in AR18 framework require Cobb-Douglas aggregation of tasks while arising endogenously in our model.

feature of the task space.

Our paper is related to the literature on the microfoundations of the Cobb-Douglas production function. The seminal work in this area is Jones (2005), whose model builds on the classic insights due to Houthakker (1955).[8] In his model a Pareto distributed menu of local production techniques featuring complementary roles of capital and labor that are widely available to firms and cost minimization gives rise to a selection of technology frontier that has the properties of Cobb-Douglas production function. While inspired by this setup and this analysis, our model fundamentally departs from it by separating capital and labor within tasks, more along the lines of AR18.

The importance of Uzawa's steady-state growth theorem and its relationship to modeling growth has been stressed by, for example, Acemoglu (2003), who provides a richer model in which capital augmenting and labor augmenting technical coexist but on the balanced growth path only labor augmenting technical arises—thus meeting the requirement of the Uzawa's theorem. Here we develop a model that does not require such a result. For an accessible version of the proof of the Uzawa's steady-state growth theorem, see Jones and Scrimgeour (2008), which is based on Schlicht (2006). Grossman et al. (2017) incorporate human capital into Uzawa's theorem and Leon-Ledesma and Satchi (2019) model the difference between short- and long-run elasticity between capital and labor.[9]

The rest of the paper is organized as follows. Section 1 presents our baseline model of production. Section 2 sets up and analyzes the leading example of the paper. Section 3 derives general conditions for balanced growth. Section 4 endogenizes capital and the notion of task complexity. Section 5 identifies and characterizes advancement to automation technology that lower the labor share. Section 6 discusses ways of endogenizing the complexity distribution. All proofs are in Appendix A at the end unless otherwise noted.

1 Baseline model of production

We begin by laying out baseline theory of production.

[8]Extended by Growiec (2013) to obtain a CES production function.

[9]In addition, our paper is also indirectly related to the literature that develops microfoundations of the R&D process, which could provide further microfoundations for the technology frontier on the task level, as we highlight in the last part of the paper. The modern probabilistic take on R&D process relies on power laws and/or extreme value distributions. This approach has been advanced by Kortum (1997), who was the first to combine the idea of a stochastic nature of research as in Evenson and Kislev (1976) with the structure of the ladders framework by Grossman and Helpman (1991). Notable advances providing further microfoundations for his line of work can be found in Ghiglino (2012)—who extends the work by Weitzman (1998) and Kortum (1997)—and the more recent work by Perla and Tonetti (2014) and Jones (2021).

1.1 Environment

The basic unit of production is a plant, which is an abstract optimizing unit representative of the economy as a whole. Plants produce a homogeneous good sold in a market for price $P > 0$. There are two factors of production: capital and labor. The user cost of capital is denoted by $r > 0$ and the wage rate is denoted by $w > 0$. Factor markets are competitive and prices are taken as given. Time is discrete and horizon is infinite, with time index denoted by $t = 1, 2, ...$ But we drop time subscript unless it is specifically is needed.

Plant technology

To produce a unit of output a plant must complete a fixed mass (measure) $M > 0$ of tasks. A task is a basic operation that can be either performed by a unit of labor, or $k(q)$ units of capital, where q represents the *complexity* that a given task presents to capital, with $q \in \mathcal{Q} = [q_0, \infty)$ and $q_0 \geq 0$.[10] The function $k(q)$ is assumed non-negative and increasing. We will refer to it as the capital requirement function.[11]

Let π be an absolutely continuous probability measure on the Borel σ-algebra $\mathscr{B}(\mathcal{Q})$ generated by $\mathcal{Q}$. Let g be the implied probability density function (pdf hereafter) and let G be its cumulative distribution function (cdf hereafter), with counter cumulative distribution function denoted by S.[12] The plant is assumed to draw the random set of tasks $\mathcal{M} \subseteq \mathcal{Q}$ from the distribution g once and for all upon its inception. Producing more output thereafter amounts to repeating these tasks again and again, resulting in a constant returns to scale technology. Summarizing, the technology of a plant comprises 3 objects $T := (\mathcal{Q}, k, (M, g))$, or interchangably $T := (\mathcal{Q}, k, (M, G))$. We denote the set of technologies consistent with these assumptions by $\mathcal{T}$. For later use, we also define the implied measure function for the task space $\mu(\mathcal{A}) = M\pi(\mathcal{A}) / \pi(\mathcal{M})$, where $\mathcal{A} \in \mathscr{B}(\mathcal{Q})$.

Plant problem

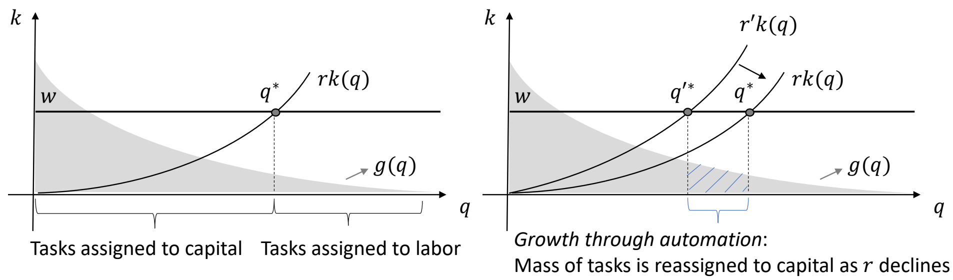
By the replication argument extended to a continuum, the plant technology is constant returns to scale.[13] As a result, the plant chooses the output level Y (scale) to maximize profits given by

[10] It is without loss to have total factor input that varies by task, for example by having a variable labor requirement $l(q)$, as long as $k(q)$ and $l(q)$ are independent across tasks. What permits this normalization is the fact that all tasks must be completed. See formal discussion in Appendix B.1.

[11] Since it is increasing, it is also continuous and differentiable except for a countable number of points.

[12] Absolute continuity of probability measure implies the existence of a probability density function.

[13] The plant can repeat all the tasks and produce twice as much output. To ease the analysis, we assume the production process is continuous and partial completion of all tasks is associated with a proportional increment of



(a) (b)

Figure 1: Cost minimization problem of a plant (a) and growth through automation (b).

Notes: Panel a illustrates the cost minimizing split of tasks to those assigned to capital and those assigned to labor. Tasks are on the horizontal axis and the cost of completing a task is on the vertical axis. The shaded area shows the distribution of tasks g. $rk(q)$ schedule shows the cost of completing a task using capital and the w schedule shows the cost of completing task using labor. All tasks with $rk(q) \leq w$ are automated (assigned to capital) and the optimum q^* occurs where the two schedules intersect. Panel b illustrates comparative statics with respect to a decline in the price of capital r, which falls from r' to r and shifts the $rk(q)$-schedule accordingly. The figure illustrates *growth through automation*, since a mass of tasks in the striped region is reassigned from labor to capital.

$\Pi = (P - c(w, r)) Y$, where $c(w, r)$ is both the unit cost and the marginal cost. For any $P \neq c(w, r)$, Y is either infinite or zero, and so equilibrium requires $P = c(w, r)$, in which case the plant chooses Y that best clears the market. We maintain the zero profit assumption throughout.

To define the cost function, we need more notation. Let $\mathscr{P}$ be a set of all measurable partitions of the task set $\mathcal{M}$ to subsets $\mathcal{M}_k$, $\mathcal{M}_l$, such that tasks in $\mathcal{M}_k$ are assigned to capital and tasks in $\mathcal{M}_l$ are assigned to labor. The plant chooses a particular member $\{\mathcal{M}_k, \mathcal{M}_l\}$ of $\mathscr{P}$ that minimizes the cost $c(w, r) := rK + wL$ subject to the technological constraints given by $L = Y \int_{\mathcal{M}_l} 1 d\mu = Y \mu(\mathcal{M}_l)$ and $K = Y \int_{\mathcal{M}_k} k(q) \, d\mu$. The first constraint states that the total labor usage L to produce Y units of output is determined by the measure of tasks assigned to labor, $\mu(\mathcal{M}_l)$. The second constraint states that the total capital usage is determined by the capital requirement function $k(q)$ integrated over the set of tasks $\mathcal{M}_k$ assigned to capital under the measure μ.

Since $k(q)$ is an increasing function, the above problem amounts to finding the *task complexity partition cutoff* q^* that divides the complexity space in such a way that all tasks below q^* are performed by capital and all tasks above q^* are performed by labor. At the cutoff point the plant must be indifferent about which factor it uses to complete this task, and so q^* solves $rk(q^*) = w$, implying

$$q^* = q_k^* \left(\frac{w}{r} \right) := k^{-1} \left(\frac{w}{r} \right). \tag{1}$$

The cutoff is unique whenever $k(q)$ is increasing in the neighborhood of such q^*.

Figure 1a illustrates the cost minimization problem. Tasks are on the horizontal axis and the cost

output.

of completing each task is on the vertical axis. The shaded area corresponds to the pdf g of tasks at each complexity level on the horizontal axis, and thus tells us about the mass of tasks at each point. As we can see, the optimum q^* lies at the intersection of the cost of performing tasks using labor (w-schedule) and the cost of performing them using capital (rk-schedule). Since q^* is a function of the factor price ratio $\frac{w}{r}$, factor intensities are a function of $\frac{w}{r}$. For later use, let us denote the optimal factor intensities by $\frac{K}{Y}\left(\frac{w}{r}\right)$ and $\frac{L}{Y}\left(\frac{w}{r}\right)$. We will refer to them as *capital intensity* and *labor intensity*, respectively. Note that production technology aggregates to a Cobb-Douglas production function whenever the factor intensities obey the identity

$$\left(\frac{K}{Y}\left(\frac{w}{r}\right)\right)^{\alpha}\left(\frac{L}{Y}\left(\frac{w}{r}\right)\right)^{1-\alpha} \equiv \frac{1}{A}, \tag{2}$$

for some $A > 0$.

Definition of *growth through automation*

We define *automation* as a displacement of labor by capital from a mass of tasks; that is, in the context of the above technology, automation pertains to an increase in the cutoff q^* as defined above (assuming $g > 0$ on that range). We define *growth through automation* as growth (steady increase in market clearing Y) driven by technical change that, on net, leads to automation at the plant level as the economy grows. Clearly, by (1), this requires an increase w/r, or an equivalent effect generated by the relative productivity of capital. The encompassing growth model will be defined in the section. For now, growth pertains to any technical change that incentivices the plant to increase output by turning its profits to positive unless P declines.

For illustration purposes, Figure 1b considers the simplest example of such a progress: an exogenous decline in the user cost of capital (from r' to r)—which can either be driven by a decline in the cost of producing capital goods or a uniform increase in the productivity of capital across all tasks. From an individual plant perspective, as illustrated in the figure, a lower r compresses the rk-schedule by moving it away from its initial position $r'k(q)$, and leads to growth through automation because such a change displaces labor from a mass of tasks corresponding the stripped area in the figure. In contrast, labor augmenting technological progress—equivalent to a growing number of effective labor units per actual worker—would have the opposite effect and lead to the displacement of capital. We do not assume away such a form of progress, but assume that the former type of progress dominates so that the economy automates as it grows; that is, we seek a result that automation is a balanced growth phenomenon as in AR18.

The question we are interested in is under what conditions on the task technology T does the plant cost minimization problem implies a Cobb-Douglas relation between the plant's total output Y and its total inputs K, L. The reason why we are interested in that condition is because this is what is required for balanced growth by Uzawa's steady-state growth theorem. But, before we get there, we discuss an analytic example highlighting how our task technology works, embed it in a on-sector growth model, and provide the key intuitions for the rest of the paper.

2 Analytic framework and leading example

Here we develop an analytic framework with explicitly defined growth model. We use this model throughout to illustrate all our general results.

Let the task technology $T \in \mathcal{T}$ of a plant be described by an isoelastic capital requirement function of the form

$$k(q) = \frac{\theta - \zeta}{\zeta} q^{\theta}/Z,$$

and let complexity be Pareto distributed with cdf

$$G(q) = 1 - \left(\frac{q_0}{q}\right)^{\zeta},$$

where $\theta > \zeta > 0$, $q \geq q_0$, and $Z > 0$ is a free constant determining the overall productivity of capital. For later use, we will set $M = \left(Aq_0^{\zeta}\right)^{-1}$, for some constant $A > 0$. We explain later why do so, and for now note that since M is a free constant and q_0, ζ are parameters, we can normalize it this way. [14]

The task complexity partition cutoff is

$$q^* = \left(Z \frac{\zeta}{\theta - \zeta} \frac{w}{r}\right)^{\frac{1}{\theta}}, \tag{3}$$

and as long as $q^* > q_0$ it falls onto the complexity domain $\mathcal{Q} = [q_0, \infty)$. Otherwise $q^* = q_0$. Consider first parameters values such that $q^* > q_0$. Since $k(q)$ is strictly increasing, we can unambiguously

[14] Calculations for this and other examples are in a Mathematica notebook available online.

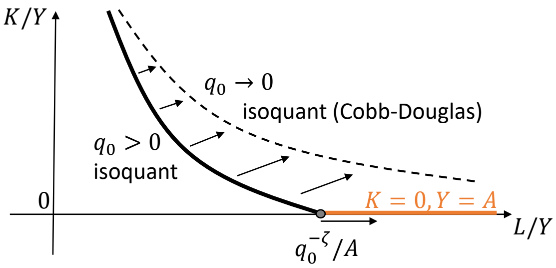


Figure 2: Representative isoquant of the plant production function in the leading example.

Notes: The figures plots the representative isoquant of the aggregate production function implied by cost minimization in the leading example. The figure shows how the isoquant behaves as $q_0 > 0$ becomes smaller and smaller, eventually uniformely converging to the Cobb-Douglas isoquant with $q_0 \to 0$, as shown in text. The isoquant has a kink at $q_0^{-\zeta}/A$, which moves to the right as $q_0 \to 0$. On the flat portion (in organge) capital is not used in production, and output is produced exclusively from labor.

obtain factor intensities, which for labor gives

$$\frac{L}{Y} \equiv M \int_{q^*}^{\infty} 1 dG = MS\left(q^*\right) = A^{-1} q^{*-\zeta} \tag{4}$$

and for capital gives

$$\frac{K}{Y} \equiv M \int_{q_0}^{q^*} k\left(q\right) dG = \left(AZ\right)^{-1} \left(q^{*\theta-\zeta} - q_0^{\theta-\zeta}\right), \tag{5}$$

after substituting for M.

Solving for q^* from (4), we obtain $q^* = \left(A\frac{L}{Y}\right)^{-\frac{1}{\zeta}}$. Plugging in to (5), we obtain the formula for the isoquant of the aggregate production function of the plant:

$$\frac{K}{Y} = \left(AZ\right)^{-1} \left(\left(A\frac{L}{Y}\right)^{1-\frac{\theta}{\zeta}} - q_0^{\theta-\zeta}\right). \tag{6}$$

It is easy to see that for $q_0 \to 0$ (since $\theta > \zeta$) the formula boils down to the Cobb-Douglas production function of the form:

$$Y\left(K, L\right) := A\left(ZK\right)^{\frac{\zeta}{\theta}} L^{1-\frac{\zeta}{\theta}}, \tag{7}$$

and so we have obtained appears to be an approximately Cobb-Douglas production function for small q_0 (we return to this below). If, on the other hand, 3 implies $q^* < q_0$, then $q^* = q_0$. In that case the domain constraint is binding and the plant only uses labor, which gives $Y_{q_0}\left(L\right) = A$, $K = 0$,and also $\frac{K}{Y} = 0$.

Figure (2) plots the obtained isoquant of the implied production function by the plant's cost minimization problem. Since technology is constant returns to scale, a single isoquant characterizes the production function, and so we refer to it as the *representative* isoquant. As we can see, the isoquant hits zero at $\frac{L}{Y} = q_0^{-\zeta}/A$ and becomes horizontal for higher values of $\frac{L}{Y}$, but below this cutoff it takes the usual shape of a hyperbola with a vertical asymptote at $\frac{L}{Y} = 0$. In particular, for any $L > q_0^{-\zeta}$, we have $Y_{q_0}(L) = A$ (and $K = 0$),and for any $L < q_0^{-\zeta}$, $Y(K, L)$ is such that it solves (6).

The important observation is that the representative isoquant converges uniformly with respect to q_0 to the Cobb-Douglas isoquant as hited before and as it is illustrated in the figure. The follows from the fact that

$$\sup_{L/Y \geq 0} |\frac{K}{Y}\left(\frac{L}{Y}; q_0\right) - \frac{K}{Y}_0\left(\frac{L}{Y}\right)| = \sup_{L/Y \geq 0} \frac{\zeta}{\theta - \zeta}\left(AZ\right)^{-1} q_0^{\theta-\zeta} \rightarrow_{q_0 \rightarrow 0} 0,$$

where $\frac{K}{Y}\left(\frac{L}{Y}; q_0\right)$ is the capital intensity for a given labor intensity as implied by (6), and $\frac{K}{Y}_0\left(\frac{L}{Y}\right)$ is the corresponding Cobb-Douglas isoquant (associated with $q_0 = 0$). Accordingly, for any sufficiently small $q_0 > 0$, our technology approximately gives a Cobb-Douglas isoquant. We will discuss next the balanced-growth path implications of this important observation when embedded into the standard neoclassical growth model.

2.1 Basic growth properties

Here we characterize growth implications of using our technology within the neoclassical growth model. The setup of the encompassing growth model is standard, but we outline it to introduce the notation.

Encompassing neoclassical growth model

Llet the aggregate feasibility be $C_t + K_{t+1} - (1 - \delta)K_t = Y_t(K_t, L_t)$, where C_t is the utility maximizing consumption of the stand-in aggregate household in period $t = 1, 2, ...$, K_t is total capital stock, and $L_t = \bar{L}$ is fixed aggregate labor supply. Capital depreciates at rate δ per period and, as in 7, the production function of the economy $Y_t(K_t, L_t)$ involves some exogenous productivity terms A_t and Z_t that grow at constant rates $\gamma_A > 1$ and $\gamma_Z > 1$, respectively.

The allocation $a := \{C_t, K_{t+1}\}_{t=0}^{\infty}$ solves the planning problem of maximizing the stand-in household's life-time utility subject to the above feasibility constraints; that is, $a = \arg\max_a \sum_t \rho^t u(C_t)$ subject to $C_t + K_{t+1} - (1 - \delta)K_t = Y_t(K_t, L_t)$, non-negativity of C_t,K_{t+1}, and given some $K_0 > 0$, where we assume $u(C_t) = \frac{C_t^{1-\sigma}}{1-\sigma}$, $\sigma > 1$. In what follows, $Y_t(K_t, L_t)$ corresponds to the implied production function of a (representative) plant in our task model. It is easy to verify that, if $Y_t(K_t, L_t)$ is given by (7), the model yields balanced growth with $t+1$ variables being multiplied by an exogenous growth factor of the form: $\bar{\gamma} := \left(\gamma_A \gamma_Z^{\zeta/\theta}\right)^{\frac{1}{1-\zeta/\theta}} > 1$ (labor supply remains fixed at $L = \bar{L}$). We makes use of this basic observation below.

Property 1: Approximate and asymptotically exact balanced growth

Consider $Y(K, L) \equiv Y_{q_0}(K, L)$ as implicitly defined by 6 for some $q_0 > 0$. Our goal is to show that this case obtains approximate and asymptotically exact balanced growth. We begin by assuming that Z is sufficiently high given q_0 (or q_0 is sufficiently small), so that capital is used in equilibrium; that is, the economy stays on the increasing portion of the isoquant in Figure 2. We return to this assumption later and show that once this is true it is always true, but for now we proceed.

For convenience, let us divide all variables by balanced-growth path factor $\left(A_t Z_t^{\frac{\zeta}{\theta}}\right)^{\frac{1}{1-\frac{\zeta}{\theta}}}$ except for labor, which is fixed at $\bar{L}$. For example, K_t becomes $\left(A_t Z_t^{\frac{\zeta}{\theta}}\right)^{\frac{1}{1-\frac{\zeta}{\theta}}} \bar{K}_t$, and so on and so forth. The planning problem after normalization implies the normalized allocation maximizes $\sum_t (\rho\bar{\gamma})^t u\left(\bar{C}_t\right)$ subject to feasibility

$$\bar{C}_t + \bar{\gamma}\bar{K}_{t+1} - (1-\delta)\bar{K}_t = \bar{Y}_{q_0,t}, \tag{8}$$

where, from (6), we have [15]

$$\bar{Y}_{q_0,t} = \left(\bar{K}_t / \bar{X}_t\left(\bar{Y}_t, \bar{L}\right)\right)^{\frac{\zeta}{\theta}} \bar{L}^{1-\frac{\zeta}{\theta}}, \tag{9}$$

and where

$$\bar{X}_t\left(\bar{Y}_t, \bar{L}_t\right) := 1 - (A_t Z_t)^{-1} \left(\frac{\bar{L}}{\bar{Y}_t}\right)^{\frac{\theta}{\zeta}-1} q_0^{\theta-\zeta}. \tag{10}$$

The first order conditions to this problem lead to the familiar Euler equation that characterizes how the economy grows from one period to the next. Had the production function been Cobb-Douglas, and had we started from the balanced-growth path value of capital $\bar{K}$, the Euler equation would hold for constant values of each variable. But, here we will evaluate it at any arbitrary level of capital $\hat{K} := \bar{K}_{t+1}$ to obtain the optimal consumption growth associated with that level of capital, which in

[15] See Appendix B.2 for an explicit derivation of this expression.

the Cobb-Douglas case gives

$$\gamma^{CD}\left(\hat{K}\right) := \rho^{-1}\left(MPK^{CD}\left(\hat{K},\bar{L}\right) + (1-\delta)\right) = \left(\frac{\bar{C}_{t+1}}{\bar{C}_t}\right)^{\sigma},$$

given

$$MPK^{CD}\left(\hat{K},\bar{L}\right) := \frac{\partial \bar{Y}_{0,t}\left(\hat{K},\bar{L}\right)}{\partial \hat{K}} = \frac{\zeta}{\theta}\left(\frac{\bar{Y}^{CD}\left(\hat{K},\bar{L}\right)}{\bar{L}}\right)^{1-\frac{\theta}{\zeta}},$$

where $\bar{Y}^{CD}\left(\hat{K},\bar{L}\right) = \hat{K}^{\frac{\zeta}{\theta}}\bar{L}^{1-\frac{\zeta}{\theta}}$ (for later use, we express MPK this way). [16]

In the approximate model, the Euler equation is the same except for the marginal product of capital that is different and time-dependent,

$$\gamma\left(\hat{K}\right) := \rho^{-1}\left(MPK_{t+1}\left(\hat{K},\bar{L}\right) + (1-\delta)\right) = \left(\frac{\bar{C}_{t+1}}{\bar{C}_t}\right)^{\sigma}.$$

We can rewrite both of the Euler equations as

$$\varepsilon\left(\hat{K}\right) := MPK_{t+1}\left(\hat{K},\bar{L}\right)^{-1} = \left(\rho\gamma\left(\hat{K}\right) - (1-\delta)\right)^{-1}.$$

which provides for a useful metric to gauge how far off the approximate model is in terms of its growth implications from the Cobb-Douglas case. Subtracting the two functions, we obtain

$$\Delta\varepsilon\left(\hat{K}\right) := \varepsilon^{CD}\left(\hat{K}\right) - \varepsilon\left(\hat{K}\right) = MPK^{CD}\left(\hat{K},\bar{L}\right)^{-1} - MPK_{t+1}\left(\hat{K},\bar{L}\right)^{-1}.$$

We will now show that for any arbitrary upper bound $\bar{\hat{K}} > 0$ we have: i) $\sup_{\bar{\hat{K}}>\hat{K}>K_0}|\Delta\varepsilon\left(\hat{K}\right)| \to_{q_0\to 0} 0$, for any t (any A_tZ_t); and ii) $\sup_{\bar{\hat{K}}>\hat{K}>K_0}|\Delta\varepsilon\left(\hat{K}\right)| \to_{t\to\infty} 0$, for any $q_0 > 0$. Given the continuity of ε with respect to γ on the restricted domain, this implies that growth is uniformly similar in two models and becomes asymptotically identical with growth. Imposing an upper bound $\bar{\hat{K}}$ is without loss by nonnegativity of $\bar{C}_t$ and (9).

The key is to calculate MPK_{t+1} by implicitly differentiating (9), which is tedious and gives

$$MPK_{t+1}\left(\hat{K},\bar{L}\right) = \left(\frac{\theta}{\zeta}\left(\frac{\bar{Y}_{t+1}\left(\hat{K},\bar{L}\right)}{\bar{L}}\right)^{\frac{\theta}{\zeta}-1} - (A_tZ_t)^{-1}q_0^{\theta-\zeta}\right)^{-1}, \tag{11}$$

[16] Note here that we are fixing $\bar{K}_{t+1}$ while assuming that this corresponds to some $\bar{K}_t$ that would lead the planner to choose this value. From this we back out growth at $\bar{K}_t$ that is then implicitly defined by this procedure.

where $\bar{Y}_{t+1}\left(\hat{K},\bar{L}\right)$ is implicitly given by (9).[17] It is clear that MPK_{CD} stated above obtains by setting $q_0 = 0$ in (11) and (10). As a result, the fact that $\bar{Y}_{t+1}\left(\hat{K},\bar{L}\right) \neq \bar{Y}^{CD}\left(\hat{K},\bar{L}\right)$, and the fact that the last term involves q_0 are the only reasons why MPK_{t+1} above might be different from MPK^{CD}.

It is ow easy to verify that

$$\Delta\varepsilon\left(\hat{K}\right) = \frac{\theta}{\zeta}\left(\frac{\bar{Y}_{t+1}\left(\hat{K},\bar{L}\right)}{\bar{L}}\right)^{\frac{\theta}{\zeta}-1}\left(\bar{X}_{t+1}\left(\bar{Y}_{t+1},\bar{L}\right)^{\frac{\zeta}{\theta}}-1\right)^{\frac{\theta}{\zeta}-1} + \left(A_{t+1}Z_{t+1}\right)^{-1} q_0^{\theta-\zeta}.$$

The last term vanishes as $A_{t+1}Z_{t+1} \to \infty$ or when $q_0 \to 0$. Furthermore, the equation for $\bar{X}_t(.)$ in (8) shows that, as $A_{t+1}Z_{t+1} \to \infty$ or $q_0 \to 0$, and also $\bar{X}_{t+1}\left(\hat{K},\bar{L}\right) \to 1$. We can take the worst case scenario; that is, use the highest value of $\bar{Y}_{t+1}$ on the stated domain $\hat{K}$ to make $\bar{X}_{t+1}\left(\hat{K},\bar{L}\right)$ as far away from 1 as possible, and pick the highest $\bar{Y}_{t+1}\left(\hat{K},\bar{L}\right)$ to amplify the implied error via the first expression. The first term corresponds to $\hat{K} = K_0 > 0$ and the second term to the upper bound $\hat{K} = \bar{\hat{K}}$. It is clear that we still obtain the convergence result, which implies convergence in the sup norm as stated. We have now shown that the representative isoquant of the plant production function is approximately Cobb-Douglas, and that growth in the encompassing one-sector growth model is approximately balanced and asymptotically balanced.

Finally, let us return to the assumption that capital is used in all periods. This trivially follows from the fact that $K = 0$ implies $\bar{Y}_{q_0,t}\left(0,\bar{L}\right) = \left(A_t Z_t\right)^{-\frac{\frac{\zeta}{\theta}}{1-\frac{\zeta}{\theta}}} \to 0$, and hence $MPK_t\left(0,\bar{L}\right) \to_{t\to\infty} \infty$. Therefore, for sufficiently high t capital will be used, and will remain in use thereafter.

Property 2: Endogenous industrial revolution and poverty traps

Before we proceed, let us briefly comment on the lower portion of the isoquant in Figure (2) and the fact that output can be produced by only labor. This feature is interesting in its own right because it leads to some interesting dynamics that can generate a stylized "industrial revolution" along the lines of Hansen and Prescott (2002) and give rise to a poverty trap.[18]

To see this, set $K_0 = 0$ and suppose growth comes exclusively from Z. If the economy does not use capital, output is $Y_t = A_t\bar{L}$. For AZ sufficiently small, it may not be optimal to accumulate capital, and hence for some time capital may not be used. Note that for $K = 0$ $MPK = \frac{\zeta}{\theta-\zeta}AZ$, and, by the

[17] See Appendix B.2 for an explicit derivation of this expression.

[18] In the sense of definition 3.2 in Azariadis and Stachurski (2005), who also provide an excellent review of the literature.

Euler equation, we must have $\frac{\zeta}{\theta-\zeta}AZ+(1-\delta)>\rho$ for capital to be accumulated ($K_{t+1}>0$). This means that capital may not be used. However, as soon as AZ exceeds a certain threshold, growth takes off with Z, starting an "industrial revolution" (use of capital in production and accelerated growth).[19]

The second feature that arises here is the possibility of a poverty trap. This arises when growth in Z comes exclusively from learning-by-doing associated with using capital. In that case, the industrial revolution may not start because the economy does not use capital at all and thus does not not "accumulate" Z. Of course, for this to work learning-by-doing must be associated with small incremental learning on the micro level that then aggregates to significant nonrival knowledge on the aggregate level, as in Romer (1986). The externality associated with not using capital then creates a free rider problem for individual agents. For more details, see Azariadis and Stachurski (2005).

Property 3: Exact balanced growth

The approximate balanced-growth path is sufficient to account for the data. In fact, amid imperfect measurement, the data only tells us that growth is approximately balanced, no more. But our model economy delivers exact balanced growth in the special case when $q_0=0$, which is an appealing result in its own right. This raises the question whether this case defines a well-defined economy on some extended domain of measures. As we show next, the answer is affirmative.

To that end, for any non-negative function $g(q)$ on $\mathcal{Q}$ and some $M>0$, let us define a set function

$$\mu_g(\mathcal{A})=M\int_{\mathcal{A}}g(q)\,dv \tag{12}$$

mapping any Borel subset $\mathcal{A}$ of $\mathcal{Q}=\mathbb{R}_+$ ($\mathscr{B}(.)$ hereafter) to a positive real number, where v is the canonical Lebesgue measure. Then, $\mu_0:=\mu_{g_0}$ for $g_0(q)=\zeta q^{-\zeta-1}$ is what we effectively integrate above as we take the limit $q_0\to 0$. As we noted earlier, this case implies Cobb-Douglas production function in (7).

A standard result in measure theory tells us that an integral of any non-negative function defines a measure, and so μ_g is a well-defined measure, just an infinite one that does not have a probabilistic representation.[20] For completeness we state this result as a proposition and the proof can be found

[19] With $K_0>0$ an analogous effect arises, but the formula for MPK is more complicated (capital will be used by deaccumulated for some time).

[20] See (Billingsley, 1995) Theorem 16.9 p.216 and the discussion of formula 16.11 on page 213. The function under the integral is trivially measurable on $\mathbb{R}_+$ as required, albeit it is not Lebesgue integrable on $\mathbb{R}_+$, which is why the resulting measure is infinite (albeit still σ-finite). Infinite measure implies $\mu_0(\mathcal{Q})=\infty$, or equivalently $\mu_0(\mathcal{M})=\infty$,

in the Online Appendix B.3. for this particular function. The only nonstandard part, however, is the most obvious one: the fact that μ_0 is a σ-finite infinite measure.

Lemma 1. $\mu_0 := \mu_{g_0}$ *is a* σ*-finite infinite measure over* $\mathscr{B}(\mathbb{R}_+)$.

There are no tractability issues associated with such more general formulation of plant technology. The integral of any measurable function $f(q)$ over such a measure has the property that $\int_{\mathcal{A}} f d\mu_g = M \int_{\mathcal{A}} f g dv$, where v is the canonical Lebesgue measure.[21] Consequently, as long as the set $\mathcal{A}$ is an interval, which is always the case in our model, the integrals convert to Riemann integrals that can be computed using standard calculus methods. For later use, let us refer to measure μ_0 as the *Pareto measure* and to the associated task complexity measure space as the *Cobb-Douglas task complexity measure space*. Furthermore, we denote the set of plant technologies with a measure generated by some density g as $\mathcal{T}^{ext} \supset \mathcal{T}$, with a generic element denoted by $T = (\mathcal{Q}, k, (M, g))$ or $T = (\mathcal{Q}, k, \mu)$. We summarize the main result we have just obtained in the proposition below. (The proof is in text above.)

Proposition 1. *Let* $T_0 = \left(\mathbb{R}_+, \frac{\theta-\zeta}{\zeta} q^\theta Z^{-1}, \left(A^{-1}, \zeta q^{-\zeta-1}\right)\right) \in \mathcal{T}^{ext}$. *The aggregate plant production function is Cobb-Douglas as in (7), and growth in the encompassing one-sector growth model is balanced.*

The Cobb-Douglas measure space is infinite and it implies that a plant needs to complete an infinite measure of tasks to produce a unit of output.[22] But this does *not* imply infinite inputs, because infinite measure of tasks derives from tasks with low q, and such tasks are also trivial for capital; that is, they use very little of it because $k(q) \to_{q \to 0} 0$. But the infinite measure is what makes capital essential for production ($K = 0$ implies $Y = 0$), and it ensures the standard Inada condition for capital. Labor is also essential because the capital requirement $k(q)$ is unbounded. Despite complexity domain is the entire real line, the finiteness of $\int_{q_0}^{\infty} g_0 dq$ ensures that labor input is finite.

Of course, one of the drawbacks of such an extended technology domain is that we can no longer think of tasks as being drawn upon the inception of the plant from some probability distribution. In this case, the task set $\mathcal{M}$ is given and it corresponds to the domain $\mathcal{Q}$, where μ_0 returns the measure of tasks of a complexity for any measurable subset. A probabilistic formulation we started

since here $\mathcal{M}$ is $\mathcal{Q}$ here.

[21] See (Billingsley, 1995) Theorem 16.11.

[22] This is not unusual, and the canonical Lebesgue measure of the real line is infinite.

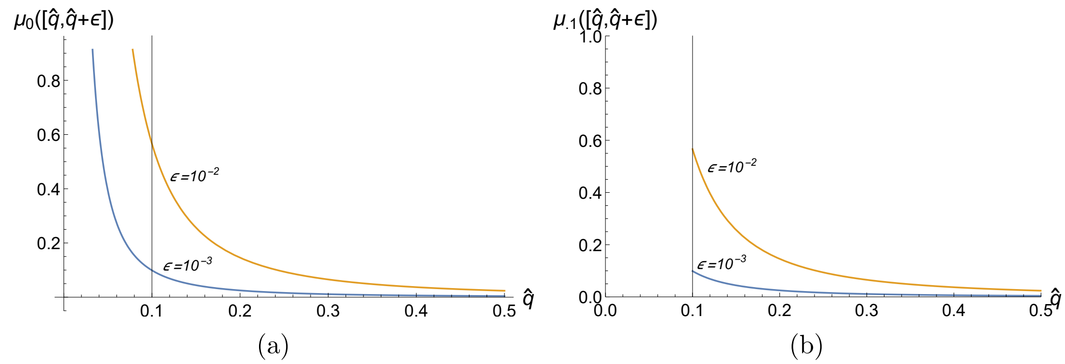


Figure 3: Properties of the Pareto measure in the leading example.

Notes: The figure plots the measure of interval $\mathcal{I} := [\hat{q}, \hat{q} + \epsilon]$ under a the Pareto measure as defined in text for $\zeta = 1$ and compares it to the implied standard Pareto probability measure for $q_0 = .1$, $\zeta = 1$ (multiplied by 10 as a unit normalization for comparison). As we can see, the two measures are identical on the part of the domain where they are both defined.

with is convenient when one wants to think about models with plant entry, endogenous evolution of plant technology, and endogenous technology adoption. For example, our model with a probabilistic representation fits into the setup a la Atkeson and Kehoe (2007). With a measure space we would need to add more structure to achieve that.

Lastly, let us illustrate the Pareto measure and compare it to the standard Pareto probability measure. Figure 3 illustrates the Pareto measure for a fixed interval $\mathcal{I} = [\hat{q}, \hat{q} + \epsilon]$ plotted with respect to $\hat{q} \geq 0$ for $\epsilon = 10^{-3}$ and $\epsilon = 10^{-2}$. The left panel shows the measure of this interval under $\mu_0(\mathcal{I})$, while the right plots the measure of the same interval using the standard Pareto probability distribution with $q_0 = .1$, which is defined analogously as in the leading example $\mu_{q_0} = M(q_0) \int_{\mathcal{I}} \zeta q_0^{\zeta} q^{-\zeta-1} dq$, with $M(q_0) = q_0^{-\zeta}$ to keep it on the same scale. As expected, the two measures coincide on the part of the domain that the latter is defined.

2.2 *Growth through automation* and factor shares

What is the mechanism that makes factor shares price-invariant in this setup? As we discuss, this mechanism has nothing to do with the infinite nature of the Cobb-Douglas measure space and pertains solely to the steepness of the capital requirement schedule.[23] This point is important to stress because, as it turns out, the requirement of a global Cobb-Douglas production function is much more stringent than the requirement that labor share does not to fall in the course of automation. The

[23]The balanced growth path consistent with Kaldor facts requires Cobb-Douglas production function because it moves the economy along the isoquant implied by the production function, given K/Y is constant by Kaldor facts, and so K/L increases because output Y grows faster than the population. It is for this reason that we will need more than local factor stability.

two properties should not be confused.

Consider the comparative statics of a decline in r as depicted in Figure 1b, and a decomposition of the factor shares implied by the zero profit condition of the plant:

$$\underbrace{\frac{w}{P}\frac{L}{Y}\left(\frac{w}{r}\right)}_{\text{labor share LS}} + \underbrace{\frac{r}{P}\frac{K}{Y}\left(\frac{w}{r}\right)}_{\text{capital share KS}} \equiv 1, \tag{13}$$

where $\frac{L}{Y}\left(\frac{w}{r}\right)$ and $\frac{K}{Y}\left(\frac{w}{r}\right)$ are optimal factor intensities defined by (4) and (5), respectively, and P is such that plant's profits are zero (this equation holds).[24] We will analyze how factor shares behave when this equation holds (P adjusts so that it holds).

As r declines, capital intensity $\frac{K}{Y}\left(\frac{w}{r}\right)$ increases and $\frac{L}{Y}\left(\frac{w}{r}\right)$ decreases, as shown in the figure. This follows from the fact that $k\left(q\right)$ is an increasing function of complexity q and labor requirement is fixed for every task. Since the wage w is unchanged, plant profits rise after the decline in r, and it follows that P must decline (restoring the identity in (13)). For the labor and capital shares to remain constant following this change, P must fall *just enough* to offset the implied decline in the labor intensity $\frac{L}{Y}\left(\frac{w}{r}\right)$—the first term in the equation above (labeled LS). Since $\Delta\frac{L}{Y} := \frac{L}{Y}\left(\frac{w}{r}\right) - \frac{L}{Y}\left(\frac{w}{r'}\right) \propto G\left(q^{*\prime}\right) - G\left(q^{*}\right)$ ("$\propto$" means *proportional* up to a positive constant of proportionality), the decline in the labor intensity is proportional to the stripped area in the figure. The size of this area, note, is independent of P, as it only depends on the ratio w/r. For the sake of the argument, suppose this decline is 1% in relative terms; that is, let $|\Delta\frac{L}{Y}/\frac{L}{Y}\left(\frac{w}{r'}\right)| = 1\%$ and suppose P falls by an offsetting 1% to keep the labor share LS constant.

For equation (13) to hold, $\frac{K}{Y}\left(\frac{w}{r}\right)$ must thus increase by $1\% + \Delta\%$, where $\Delta\% > 0$ is how much r decreased in relative terms. Again, this is mechanically implied by the formula above assuming the labor share does not change, which implies that the complementary capital share cannot change. The conditions for this turns out key to understand what preserves factor shares in this model.

The analysis of how factor intensities are obtained reveals that a decline in $\frac{K}{Y}\left(\frac{w}{r}\right)$ can only be larger in relative terms than the decline in $\frac{L}{Y}\left(\frac{w}{r}\right)$ when the $k\left(q\right)$ schedule is sufficiently steep over the entire region of tasks where capital is used (starting from $q_0 = \inf \mathcal{Q}$). The reason is that the steepness of this schedule is what determines the initial level of $\frac{K}{Y}\left(\frac{w}{r'}\right)$ relative to its change on

[24] Note that this equation can be equivalently written as $PY = wL + rK$, which is the zero profit condition of the plant. While cost minimization alone implies a Cobb-Douglas aggregate relationship between total output and total inputs, this does not mean that the labor share equals to its exponent without imposing the zero profit condition. This is an equilibrium statement even in the standard case because the condition $MPL = w/P$ from which it is derived is ill-defined without it. Put differently, cost minimization, which is defined off equilibrium, is not sufficient to establish that the labor share is the exponent. It only establishes that the labor share in cost is the exponent.

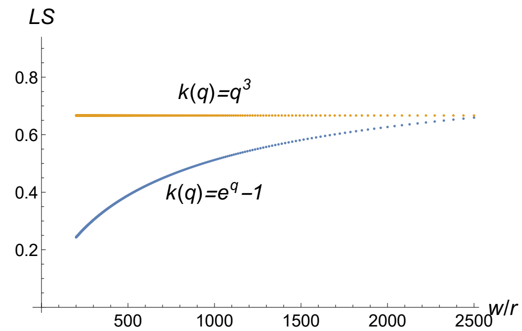


Figure 4: Steepness of $k(q)$ schedule and the labor share over the course of growth through automation.

Notes: The figure plots the labor share in two cases. The first case is the leading example with G being Pareto and assuming $\theta = 3$, $\zeta = 1$, $A = 10$, $Z = 1$ and $q_0 = .1$. The second case replaces capital requirement function by $k(q) = e^1 - 1$. As we can see, growth trough automation results in an increase in the labor share under this specification (growth through automation corresponds to an increase in the w/r ratio).

the margin: $\Delta\frac{K}{Y} := \frac{K}{Y}\left(\frac{w}{r}\right) - \frac{K}{Y}\left(\frac{w}{r'}\right)$, which is implied by the integration of $k(q)$ schedule over the (marginal) stripped area.

Concluding, this shows that if the advantage of humans over machines is slim across vast swathes of tasks, as implied by a flat $k(q)$ schedule relative to $g(q)$, the labor share will fall in the course of automation. But as long as labor's advantage is large—corresponding to a steep $k(q)$ schedule relative to $g(q)$—this can be enough of a defense line to prevent automation from reducing the labor share, or even result in its increase. We refer to this effect as the effect of *diminishing returns from automation* associated with rising task complexity.

As we stressed in the beginning, this property has nothing to do with the fact that the measure μ_0 is infinite, and thus whether technology has a probabilistic representation or not. We illustrate this important observation with a concrete numerical example depicted in Figure (4).

The capital requirement function in this case is $k(q) = e^q - 1$,and we let complexity q be distributed according to a Pareto distribution with pdf $g(q) = \zeta q_0^{\zeta} q^{-\zeta-1}$ and $q_0 > 0$. The capital requirement function is globally steeper than, for example, $k(q) = q^3$, which would correspond to the setup in the leading example with $\theta = 3$. We include it for comparison. We solve both cases numerically for $\zeta = 1$, $q_0 = .1$, and plot the result for a range of values for w/r. As we can see, when $k(q)$ is steeper, as implied by $k(q) = e^q - 1$ versus $k(q) = q^3$, growth through automation *increases* the labor share.

Finally, while we focused here on the labor share, it is important to note that labor may benefit from automation even when the labor share declines. This is because the price P of a good falls due to an initial increase in profits, which raises the real wage $\frac{w}{P}$ (ceteris paribus). Our theory distinguishes between these effects and connects them in a particular way, but since this is not our focus we leave the analysis of this property to the interested reader.

Human capital and automation

We do not incorporate human capital but our framework can also speak to the division of labor income between skilled and unskilled labor. For example, consider $h(q) = q^{\chi}$, where $0 < \chi < 1$, and suppose labor input to complete a task must come in a bundle of 1 unit of unskilled labor that costs w and $h(q)$ units of human capital that costs w^h. Such an extension readily follows from our analysis and requires adding human capital as an additional factor. Automation will also increase human capital's share in labor income. The production function will remain Cobb-Douglas between capital and labor, but part of the income will go to human capital.

3 General conditions for balanced growth

We now turn to derivation of general conditions that ensure balanced-growth path with automation. By the Uzawa's theorem, we know this requires the production function must be Cobb-Douglas.[25] We have shown that such an outcome is possible. But what is the set of technologies with such a property? Is this example the only one possible or it is a member of larger class? Knowing the answer is important because it may provide a basis for what is empirically plausible in terms of functional forms without digging into the details of the makings of the task based production technology on the micro level.

The proposition below states our key result. It shows that the production function is Cobb-Douglas *if and only if* the listed conditions hold. For now, they are abstract, but the subsequent discussion will narrow down on their implications. (Throughout the rest of the analysis the symbol C serves as a generic and context dependent constant. Its occurrence in various unrelated contexts does not imply it is the same constant. We use $\hat{C}$ to distinguish constants when needed.)

Proposition 2. *The aggregate production function implied by the cost minimization problem of the plant is Cobb-Douglas with exponent* $0 < \alpha < 1$ *iff task technology* $T = (\mathcal{Q}, k, (M, g)) \in \mathcal{T}^{ext}$ *satisfies:*

1. *Density g is admissible: a) g has full support on $\mathcal{Q}$; b) the survival function $S(q) := \int_{[q,\infty]} g dv < \infty$ is well-defined for all $q \in \mathcal{Q}$, and c) the hazard rate of complexity $h := \frac{g(q)}{S(q)} < \infty$ is well-defined for all $q \in \mathcal{Q}$.*

[25] We refer the reader to Jones and Scrimgeour (2008) for an accessible and intuitive proof of this fact.

2. *Capital requirement $k(q)$ is α-hazard monotone; that is*

$$\alpha \frac{k'(q)}{k(q)} = h(q) := \frac{g(q)}{S(q)}, \tag{14}$$

where $S(.)$ is as defined in 1 above.

3. $k(q) \rightarrow_{q \rightarrow q_0} 0$, *where $q_0 := \inf \mathcal{Q}$, and $k(q)\, g(q)$ in Lebesgue integrable on $\mathcal{Q}$.*

The proposition tells us that for a well-defined density g (under condition 1), the Cobb-Douglas function obtains when g and k are in a particular relationship described by conditions 2 and 3 above. Condition 2 states that the steepness of capital requirement $k(q)$ must be given by the hazard rate implied by density g, with the constant of proportionality corresponding to the exponent of the Cobb-Douglas production function. Condition 3 imposes $k(q_0) = 0$, which selects a particular solution of the ODE in (14). Note that the proposition stated for the general domain $\mathcal{T}^{ext}$.

It is easy to see that (14) implies that condition 3 demands the hazard rate converges to zero at $q = q_0$, which given the fact that $S(q)$ is a decreasing function, implies $S(q)$ must shoot up to infinity at $q = q_0$. What this means is that the Cobb-Douglas production function demands an infinite measure space. This important impossibility result tells us that the limit we considered in our analytic framework was essential to obtain the exact result. We summarize this result in the corollary below, after establishing the solution of the differential equation in (14) in the lemma below.

Lemma 2. *Conditions 2 and 3 in Proposition (2) imply that for any $\varepsilon > 0$ and all $q \geq q_0 + \varepsilon$, we must have*

$$k(q) = c(\varepsilon)\, CS(\varepsilon)^{\frac{1}{\alpha}}\, S(q)^{-\frac{1}{\alpha}}, \tag{15}$$

for some some fixed constant $C > 0$, and some positive valued function $c : \mathbb{R}_+ \rightarrow \mathbb{R}_{++}$ such that $c(\varepsilon) \rightarrow_{\varepsilon \rightarrow 0} 0$.

Corollary 1. *(Impossibility of balanced growth under finite measures) If g is a probability density, Lemma (2) implies*

$$k(q) = CS(q)^{\frac{1}{\alpha}} \tag{16}$$

and condition 3 implies $C = 0$, which contradicts $T \in \mathcal{T}$; hence, $T \in \mathcal{T}^{ext}/\mathcal{T}$.

Without knowing k or g, it is not possible to say more in the general case pertaining to $\mathcal{T}^{ext}$. The proposition tells us that given g the function k can be defined to ensure Cobb-Douglas aggregation, and all the around, but it does not tell us what these functions are individually. However, by requiring an approximation for the balanced growth with a finite measure space (g is probability density), as in our analytic framework, we can obtain an independent restriction for g. It turns out that analytic setup is the only possibility (with a single degree of freedom). We explore it next.

3.1 Approximate balanced growth and Pareto distribution

Consider now a probabilistic setting in which g is a probability density and the survival function is $S(q) = 1 - G(q)$, where $G(q)$ is the cumulative distribution function implied by g. Let the total measure of tasks be finite, $0 < M = A^{-1} < \infty$, as required by this setup. We already know that there is no exact balanced growth, but we also know that the Pareto distribution provides a good approximation when extended to the entire real line. There are significant advantages associated with a probabilistic setup, as discussed, which leads to the question whether our analytic setup (leading example) is the only one possible.

The next corollary states the first part of our results: All probability densities g on a the same domain Q imply exactly the same isoquant of the aggregate production function after imposing conditions 1 and 2 of Proposition (2). The key intuition is that such approximations drop the requirement that the constant associated with the solution of ODE in (14) is zero, which is key to Corollary (1). But since the constant of the ODE has nothing to do with the assumed distribution g, the result then follows, as it is implied by that constant not being zero. (The proof has been relegated to Online Appendix B.4 since it readily follows from the proof of the proposition.)

Corollary 2. *Let $T_1 = (Q, k_1, (M, G_1)) \in \mathcal{T}$ and $T_2 = (Q, k_2, (M, G_2)) \in \mathcal{T}$ be two technologies such that $k_i = CS_i^{\frac{1}{\alpha}}$, $i = 1, 2$, satisfies equation (16) for a common constant $C > 0$. Then, their implied isoquants are identical; that is, $\frac{K}{Y}_1\left(\frac{L}{Y}\right) = \frac{K}{Y}_2\left(\frac{L}{Y}\right)$. Furthermore, for $w \to \infty$ (fixed $r > 0$), if $k_i(q)\, g_i(q)$ is bounded away from zero for a sufficiently large q, the factor share ratio $\left(w\frac{L}{Y}_i\left(\frac{L}{Y}\right)\right) / \left(r\frac{K}{Y}_i\left(\frac{L}{Y}\right)\right)$ converges to $\frac{1-\alpha}{\alpha}$.*

It is best to see this result in action. To that end, consider an exponential probability distribution $g = \lambda e^{-\lambda q}$, $\mathcal{Q} = [0, \infty)$, and some positive mass $M = A^{-1} > 0$ of tasks drawn from this distribution. Let k be consistent with condition 2 of Proposition 2, which implies $k(q) = Ce^{\frac{\lambda}{\alpha}q}$ by (15), and let us replace C by $Z^{-1}\frac{1-\alpha}{\alpha}$ for the sake of a cleaner exposition.

Following the same derivation as in our leading example in Section 2, it is easy to verify that the representative isoquant of the aggregate production function is

$$\frac{K}{Y} \equiv (AZ)^{-1}\left(\left(A\frac{L}{Y}\right)^{1-\frac{1}{\alpha}}-1\right). \tag{17}$$

The functional form is identical to (6), except that q_0 is replaced by "1," which we no longer can take eliminate. But this result raises a question whether can still use this function as an approximation for Cobb-Douglas? Superficially, setting Z sufficiently large, and A sufficiently small, will make the last term "1" unimportant, implying an approximately Cobb-Douglas isoquant.

The problem with this approach is that we cannot start from a good approximation without assuming a high productivity of capital from the get go. While in the leading example we used a normalization, it was done via TFP term A (total measure of tasks M) rather than Z, which makes a difference because the former can be thought of as a change of unit in the model, while the latter cannot. In particular, Z affects the relative factor price $\frac{w}{r}$. The last part of the above corollary tells us why this actually works. As $\frac{w}{r} \to \infty$, implying $\frac{L}{Y} \to 0$, the isoquants of all distributions converge to the Cobb-Douglas isoquant, and that is what we would be relying upon. But the problem with this approach is that at that point automation apocalypse has already happened and wiped out all labor income that it could wipe out. It is easy to verify that the labor share declines from 1 to $1-\alpha$ as we move up along the isoquant, which we omit for brevity.

We conclude by showing that the approximating probability distribution has to be Pareto and the approximation must involve the lower bound of the domain q_0. This is the only way around the impossible condition implied by (15) in Lemma (2). Specifically, under Pareto probability distribution, we have $k(q) = c(\varepsilon)\, C\left(q_0^{\zeta}\varepsilon^{-\zeta}\right)^{\frac{1}{\alpha}} S\left(q_0^{\zeta}q^{-\zeta}\right)^{-\frac{1}{\alpha}}$, and while q_0 cancels out, ε does not. Importantly, we can define $c(\varepsilon) = \varepsilon^{\zeta}$, which applies for $\varepsilon \geq q_0$. We can satisfy the requirements of the lemma by taking the limit $q_0 \to 0$.

It turns out this result is general and only the Pareto probability distribution is capable of such a "trick." Unfortunately, the formal result is not particularly illuminating. The key property is the observation that it must be possible to factor out a term involving the lower bound of distribution q_0 from the survival function $S(q) = 1 - G(q)$, as only then the remainder can shoot up to infinity, which is also part of the proof. The proof then shows that imposing such requirement implies a functional form for $S(q)$ akin to Pareto distribution as technology converges to Cobb-Douglas case (up to a monotone transformation f as stated below).

Definition 1. *A cumulative distribution function G is $f(q)$-**warped Pareto** if f a positive-valued, strictly increasing, and absolutely continuous function $f:[q_0,\infty)\to R_{++}$with $f(q)\to_{q\to 0}0$ and $G(q;q_0)=1-\left(\frac{f(q_0)}{f(q)}\right)^{\zeta}$, where $q\in[f(q_0),\infty)$, $\zeta>0$.*

Remark 1. Standard Pareto distribution is a warped Pareto distribution for $f(q)=q$.

Proposition 3. *Suppose that there exists a sequence of parameterized plant technologies $T_n=(\mathcal{Q}_n=[q_{0,n},\infty),k,(M_n,g_n))\in\mathcal{T}$, $n\in\mathbb{N}$, satisfying condition 1 of Proposition (2). Furthermore, suppose T_n converges to $T_\infty\in\mathcal{T}^{ext}$ in the following sense: 1) $q_{0,n}$ converges to $q_0\geq 0$, and 2) $\int_{\mathcal{A}\cap\mathcal{Q}_n}fg_n dv$ converges to $\int_{\mathcal{A}}fg_n dv$ for any $\mathcal{A}\in\mathscr{B}(\mathcal{Q}_\infty)$ and any non-negative continuous and bounded function f on $\mathcal{A}$.[26] Then, if T_∞ aggregates to a Cobb-Douglas production function, except for finitely many terms, (M_n,g_n) is a sequence of f-warped Pareto distributions, for some function f, and $f(q_{0,n})\to_{n\to\infty}0$.*

4 Capital as a long-lasting embodiment of tasks

In our model tasks are a primitive step with respect to labor but not with respect to capital, and hence the notion of task complexity pertaining to capital. This is a natural assumption because producing a piece of capital suitable to complete a task is a different undertaking from using labor to complete the task directly. For example, while it may be practical to think of the task of driving as a primitive task with respect to labor, to produce a self-driving car one needs to think about a number of unrelated tasks that produce a self-driving module, such as obtaining an appropriate sensor suite, computing modules, and an AI engine that recognizes objects and "comprehends" the notion of driving.

Our growth model already imposes such a structure on average, by implying that producing a unit of capital is the same as producing a unit of aggregate good. But such a model is not entirely

[26] *This notion of convergence is a version of weak convergence of the (implied) measure $\mu_n(\mathcal{A}):=\int_{\mathcal{A}}g_n dv$ to some measure $\mu_\infty(\mathcal{A}):=\int_{\mathcal{A}}g dv$, whose existence and structure we impose as a condition of convergence. But the difference is the potentially changing domain. To see how it relates to weak convergence of measures, suppose $\mathcal{Q}_{n+1}\supseteq\mathcal{Q}_n$, as in the leading example. This is the only case that creates an issue, since the opposite is just weak convergence. Note that, in that case, the set $\mathcal{A}\in\mathscr{B}(\mathcal{Q}_\infty)$ corresponds to the limit $\cup_{i=1}^n\mathcal{A}_n\nearrow\mathcal{A}$, where $\mathcal{A}_n=\mathcal{A}\cap\mathcal{Q}_n$. On each set $\mathcal{A}_n$ weak convergence of measure would imply $\int_{\mathcal{A}_m}fd\mu_n\to_{n\to\infty}\int_{\mathcal{A}_m}fd\mu_\infty$ (*), for any f continuous and bounded. By monotone convergence theorem we automatically have $\int_{\mathcal{A}_n}fd\mu=\int_{\mathcal{A}}1_{\mathcal{A}_n}fd\mu\to_{n\to\infty}\int_{\mathcal{A}}fd\mu$, but to obtain the double limit $\int_{\mathcal{A}_n}fd\mu_n\to_{n\to\infty}\int_{\mathcal{A}}fd\mu$ we additionally need (*) to converge uniformly by the double limit theorem (Moore-Osgood theorem). The requirement of f being bounded is sufficient because for any finite w/r we will either require that $k(q)$ is bounded or restrict attention to bounded domain on which $k(q)$ is actually used for that price. Labor requirement is 1 and it is a bounded function.*

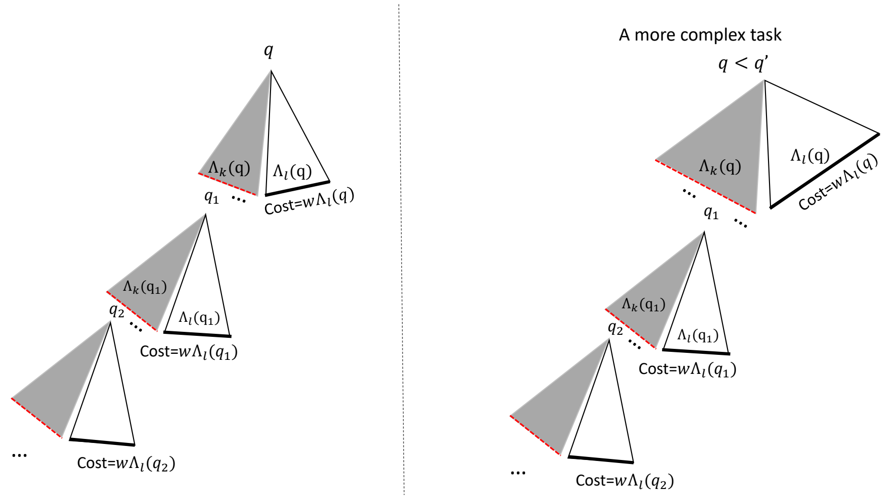


Figure 5: Task map of capital of type q (left panel) and $q' > q$ (right panel).

Notes: The figure shows a task map of capital (left panel). As shown, it is a directed graph of tasks embodied by a unit of capital of type q under cost minimization. Task map traces all tasks back to labor. The range of tasks at the top denoted by the dotted red line "q_1..." and solid black line represent the set of capital-defining tasks, which, if completed by labor, obtain a unit of capital. However, under cost minimization, the set Λ_k is best completed by capital. This expands this portion downwards to trace it back to labor (rays along the grey triangles). The comparison of the left- and right-panels illustratesthe the baseline definition of complexity as a larger measure of tasks with on average the same composition. Since the composition of tasks is on average, the rest of the tree is identical across the two figures.

satisfactory because it hides from us the underlying theory of "α," as it boils down to how capital is produced to determine its cost. Our goal is to identify and characterize technologies that may affect it, and so we need an explicit theory. As we will see, the next section will take advantage of this more general framework.

4.1 Production of capital

We define capital of type q as a lasting embodiment of some set $\Lambda(q)$ of tasks, which, if performed once, can be used repeatedly in the form of matter of sort to complete task q repeatedly—say, up to a Poisson arrival rate δ with which capital disintegrates after each use. Let $\Lambda(q) \subset \mathcal{Q}$ be the set of *type q capital-defining tasks.* We will assume that on the domain $\Lambda(q)$ density $\tilde{g}_q$ describes complexity in this set and generates some measure of tasks $\tilde{\mu}_q$. As we show, $k(q)$ in this world pertains to the price of a piece of capital to perform task q, which we assume is its resource cost, and r is then the user cost of capital normalized by this price (as we will derive).

Cost minimization implies that some capital defining tasks $\Lambda(q)$ are optimally completed by capital and some are optimally completed by labor. Accordingly, for any factor price ratio $\frac{w}{r} > 0$, there is a partition of the capital-defining set of tasks $\Lambda(q)$ to sets $\Lambda_k(q)$ and $\Lambda_l(q)$ such that tasks in $\Lambda_k(q)$ are optimally completed using capital and tasks in $\Lambda_l(q)$ are optimally completed by labor.

Formally, for any q, $q' \in \Lambda_k(q)$ if $k(q) > \frac{w}{r}$.[27] For later use, let us define the measure of tasks that are optimally completed by labor as $\lambda_q := \mu_q(\Lambda_l(q))$ which we additionally assume is a finite number for all q.

The task map of capital

Since capital may be produced by capital, its production involves a fixed point. Solving this fixed point boils down to uncovering the tasks that capital used to produce capital entails, and the costs associated with these tasks. As illustrated in Figure 5 (left panel), this reasoning leads to the notion of the *task map of capital.* We will need to make some assumptions regarding its structure to recursively define the resource cost of obtaining a unit of capital of type q.

The task map is a directed graph with an ever expanding tree along the tasks optimally assigned to capital at each stage, as seen in the figure. For example, the task of building capital of type q (at the top left) expands to the set of tasks completed by labor (solid line projected along the white rays) and capital (dotted line projected along the shaded rays). Using the same definition, tasks on the dotted line are possible to obtain from capital and labor in a cost minimizing way. This leads to the analogous split at the lower level of subordination of tasks assigned to capital at the higher level, and so on and so forth. The figure shows an example of one such task, "q_2," but an analogous expansion applies to all tasks on the dotted continuum of tasks assigned to capital at the parent level.

To formalize this idea, a natural assumption is to have sets $\Lambda(q)$, $\Lambda_k(q)$ and $\Lambda_l(q)$ apply regardless of the level where they appear. For example, if a recipe to produce a truck requires to produce a diesel engine as part of it, it is natural to think of the production of the engine within a larger task of building a truck as being the same as the task that pertains to the engine itself (a standalone entity or part of another product). With this definition in hand, the fixed point defining the price of capital as its resource cost is recursively given by law

$$k(q) := \int\limits_{q' \in \Lambda_k(q)} rk(q')\,\tilde{g}_q dv + w\lambda_q, \tag{18}$$

where $k(q)$ on the left is the cost of obtaining a unit of type q capital in units of P and it also appears on the right-hand side by the above assumption. By definition, tasks assigned to capital (the first term on the right) are in set $\Lambda_k(q)$ and their cost is $rk(q)$. We integrate this function over density

[27] The sets Λ_k, Λ_l are measurable because k is increasing and Λ is measurable.

$\tilde{g}_q$ which now is q specific. Tasks assigned to labor are treated analogously but since the labor input is 1 this part corresponds to just the measure of these tasks times the wage rate.[28] Recall that we assumed $\tilde{g}_q$ is such that λ_q is finite.

User cost of capital redefined

Before we proceed, we redefine the user cost of capital.

Let $R(q)$ be the user cost of using capital of type q, and let is be associated with the dividend (profit) earned on having this piece of capital for one period and renting it out. The usual derivation involves a zero profit condition on this dividend, which we revisit here for clarity. The zero profit condition states that $R(q)$ is such that given the costs of obtaining the piece capital for a duration of one period one would break even.

Let t denote the time period. The cost of obtaining capital and holding it for one period is its purchase price at time t, which is $k_t(q)$, the opportunity cost of funds $\rho k_t(q)$ incurred over that period (ρ is the interest rate). The resell price after that period is $k_{t+1}(q)$, but since the resell takes place with probability $1-\delta$ (with probability δ capital disintegrates), the expected residual value is $(1-\delta)\, k_{t+1}(q)$. The zero profit condition requires

$$\underbrace{R_t(q)}_{\text{user cost}} = \underbrace{(1+\rho)\, k_t(q)}_{\text{acquisition cost}} - \underbrace{(1-\delta)\, k_{t+1}(q)}_{\text{residual value after a period of use}},$$

which defines the user cost of capital.

Assuming balanced growth, let us suppose that k grows at some fixed rate $\bar{\gamma} > 1$ from one period to the next; that is, $k_{t+1}(q) = \bar{\gamma} k_t(q)$. This further simplifies the above expression to $R_t(q) = (1+\rho-(1-\delta)\bar{\gamma})\, k_t(q)$. Given how we used r in the previous section, it is clear that it corresponds to $r = 1+\rho-(1-\delta)\bar{\gamma}$. As we can see, r is independent of P, since P affects the units in which k is measured and k is the price of a piece of capital.[29] Of course, we can always return to the usual definition by defining real capital as $\hat{k}_t(q) = k_t(q)/P_t$, and use $\hat{r} = P_t r$ instead, but we will not do so. Instead, from now on, we assume output is the numeraire by setting $P = 1$. We will continue to refer to r as the user cost of capital.

[28] As a side note, observe that our task-based notion of capital overcomes the neoclassical theory's conundrum that $K = 0$ implies $Y = 0$. Since capital here is ultimately derived from labor the first unit can always be produced.

[29] It is not independent of its evolution that may influence ρ; that is, the inflation rate P_{t+1}/P_t.

4.2 Task complexity as a measure of capital-defining tasks

We now propose an explicit notion of task complexity to link density g to production of capital. In principle, there are two aspects of the above task-based notion of capital that may give rise to complexity.

The first notion of pertains to the measure of capital-defining tasks $\Lambda(q)$ that may be different. Specifically, it would make sense to think that more complex tasks require a larger measure of tasks to be completed; that is, $\mu_{q'}^{\Lambda} > \mu_q^{\Lambda}$ whenever $q' > q$. For example, producing an airliner may require completing more tasks than producing an agricultural tractor, and for that precise reason the task of *flying passengers* associated with the airliner is more complex for capital than the task of *cultivating land* associated with the tractor.

Alternatively, complexity may project downward onto the subordinate tasks along the task map (Figure 5). Intuitively, individual tasks involved in building an airliner may be on average more complex themselves because building a jet is more complex, and so capital pieces that are used to make it are more complex too.

In what follows, we assume away downward projection of complexity. Conceptually, we see this a superior approach, reflecting the fact that the task space has been properly partitioned to operations that are basic enough so that they are "on average" homogeneous. Nonetheless, the Online Appendix B.6 studies the possibility of the downward projection of complexity and provides an analytic reasons why this simple structure is more natural. There, we show that downward project of complexity leads to non-generic assumptions to obtain a Cobb-Douglas production function.

Definition of complexity

Let *complexity* q be related to the measure of capital-defining tasks; that is, a task q is more complex for capital because the associated relative measure μ_q^{Λ} is bigger. But assume that the composition of tasks does not change (as discussed below). This notion is as illustrated in Figure 5. Higher q' involves a bigger set of tasks at the first level of subordination, but since each task within this set is on average the same as for q, on average, the rest of the tree is identical.

A simple way to say that "the composition of tasks" does not change with q would be to assume that μ_q^{Λ} varies and assume $\tilde{g}_q = \mu_q^{\Lambda}\tilde{g}$,where $\tilde{g}$ is some base density (resulting in finite or infinite measure). However, in a measure-theoretic setting, this turns out to be a restrictive definition for reasons that are technical and not necessarily meaningful in terms of the economics of the problem.[30] To avoid such a strict definition, we assume that density does not vary *systematically* with q on any of the $\Lambda(q)$ sets. The assumptions we impose from now on are as follows:

Assumption 1. *For all $q \in \mathcal{Q}$, let $\mu_q^\Lambda > 0$ be a finite-valued function such that for any $q < q'$, $\mu_q^\Lambda < \mu_{q'}^\Lambda$. Assume the density of complexity $\tilde{g}_q$ on each set $\Lambda(q)$ is such that $\tilde{g}_q = \mu_q^\Lambda \tilde{g}^0_{i(q)}$, where $\tilde{g}^0_{i(q)}$ is a countable family of density functions defined on $\mathcal{Q}$, where $i(.)$ is an i.i.d. step function such that on any closed nonempty and non-degenerate interval $\mathcal{I} \subseteq \mathcal{Q}$ $\tilde{g}(q') := \mathbb{E}\left(\tilde{g}^0_{i(q)}(q') \,|q \in \mathcal{I}\right)$ and $\mathbb{E}(i(q)\,|q \in \mathcal{A}) = \mathbb{E}(i(q))$, where $\mathbb{E}$ is defined under the probability measure induced by the canonical Lebesgue measure on $\mathcal{I}$. Furthermore, assume $\tilde{g}_{i(q)}$ is such that λ_q is finite.*[31]

Before we derive our general result, we extend our leading example from Section 2 and show how this notion of complexity works. As before, this example will be helpful to understand the rest of this section. Not surprisingly, we will obtain Cobb-Douglas production function, but the key here is that we do it explicitly. (The Online Appendix B.5 provides an explicit solution of the fixed point implied by (18) under Assumption 1. Below we only use this equation in its recursive form.)

Leading example continued

Consider the setup from the leading example in Section 2. In addition, let the measure of capital-defining tasks be $\mu_q^\Lambda = (AZ)^{-1} q^{\frac{1}{\alpha}}$, and assume the distribution of capital-defining tasks at any complexity q is the same density as the one in the production of goods (defined on $\mathcal{Q} = [0, \infty]$). We will show that we do not need to specify the $k(q)$ function and we will obtain Cobb-Douglas aggregation for both sectors.

Consider first equation (18), and note that in this example it takes a simple form

$$k(q) := (AZ)^{-1} q^{\frac{1}{\alpha}} \left(\int_0^{q^*} rk(q)\, g(q)\, dq + wS\left(q^*\left(\frac{w}{r}\right)\right)\right). \tag{19}$$

Since in (18) we integrate over $\Lambda(q)$, the implied measure $(AZ)^{-1} q^{\frac{1}{\alpha}}$ of the capital producing set of

[30]There does not exist a partition of the real line into two measurable sets that split every single interval in the exact same proportion in terms of the Lebesgue measure of the two pieces (that is, their intersection with the interval). This is a consequence of the Lebesgue density theorem, which states that the measure is "dense," and our measure is likely to inherit this property. As a result, we would be implying that the sets $\Lambda(q)$ and $\Lambda(q')$ are simple replicas of the same set of tasks (characterized by some common distribution function and a varying mass of elements drawn from this distribution).

[31]A step function implies that it changes value countably many times on the domain and each subsequent value is i.i.d.

tasks appears here as a scaling factor in front of the expression on the right-hand side.

Recall from the plant cost minimization that $q^*\left(\frac{w}{r}\right) = k^{-1}\left(\frac{w}{r}\right)$, and so

$$S\left(q^*\left(\frac{w}{r}\right)\right) = \left(k^{-1}\left(\frac{w}{r}\right)\right)^{-\zeta}.$$

For now, let us assume k is such that its inverse is well-defined (we will return to this).

Next, from the proof of Proposition 2, observe that that Cobb Douglas aggregation obtains if and only if

$$\int_0^{q^*} rk(q)\, g(q)\, dq \equiv \frac{\alpha}{1-\alpha} rk(q^*)\, S(q^*)\,.$$

To see this result, associate output Y with the number of units of capital that are produced in the capital sector q and K, L with the total factor usage. The first part of the proof of Proposition 2 (Part I) applies without any changes and shows this condition is indeed equivalent to requiring Cobb-Douglas production function in the production of capital. We will use this equation to replace the integral on the left-hand side of (19).

Substituting the above two expressions in (19) we have

$$k(q) = (AZ)^{-1} q^{\frac{1}{\alpha}} \left(k^{-1}\left(\frac{w}{r}\right)\right)^{-\zeta} \frac{w}{1-\alpha}.$$

As we can see, capital requirement function follows and it is almost of the desired form. However, it is implicitly defined because it involves its own inverse. We do not need an explicit form to obtain Cobb-Douglas aggregation, since $k^{-1}\left(\frac{w}{r}\right)$ is just a number, but we do need to have a good understanding of this endogenous term (and know it is well-defined).

It turn out that this term is no more but the unit cost of production $c(w, r)$, since

$$\kappa(w, r) := A^{-1}\left(k^{-1}\left(\frac{w}{r}\right)\right)^{-\zeta} \frac{w}{1-\alpha} = A^{-1} S\left(q^*\left(\frac{w}{r}\right)\right) \frac{w}{1-\alpha},$$

To see this, recall that $L = A^{-1} S\left(q^*\left(\frac{w}{r}\right)\right)$ and so the first term in the last expression is the total labor input in production of capital, while $1-\alpha$ is the targeted labor share in value added. Since we can rewrite the above equation as $\kappa(w, r)(1-\alpha) = Lw$, and on the right we obtain unit labor cost, i\it must be that $\kappa(w, r)$ is the total unit cost. Since we set $P = 1$, we have $\kappa(w, r) = 1$.[32]

[32]We can also solve this function explicitly, obtaining $k(q) = Z^{-1} q^{\theta}\left((Z^{\alpha} A_0)^{-1} r^{\alpha} w^{1-\alpha}\right)^{\frac{1}{1-\alpha}}$.

After we set $\alpha = \frac{1}{\theta}$, capital requirement function is qualitatively the same as the one we had earlier: $k(q) = Z^{-1} q^{\frac{1}{\alpha}} \kappa(w, r) = Z^{-1} q^{\frac{1}{\alpha}}$. Since Z is an arbitrary constant, with an abuse of notation, we will scale it up to fit our leading example exactly and use instead $k(q) = Z^{-1} q^{\frac{1}{\alpha} \frac{\theta-\zeta}{\zeta}}$. It is clear that this is without loss.

We are now ready to derive the aggregate production function for both the capital q producing sectors (all identical) and the goods producing sector. As for the capital producing sector, Y corresponds to the number of units produced by the type q capital sector, and K, L measure total factor inputs. Following the same steps as in the leading example, we obtain[33]

$$Y = A (ZK)^{\frac{\zeta}{\theta}} L^{1-\frac{\zeta}{\theta}}. \tag{20}$$

Generalization

Balanced growth with capital producing sector requires the production function in that sector to be Cobb-Douglas. To obtain our general result, we follow the logic of the example above and impose this requirement in abstract. Since there may be an i.i.d. variation across q, we define what we mean by the requirement that production function is Cobb-Douglas in the definition below. Then, as the proposition below shows, if we assume the measure of capital-defining is $\mu_q^\Lambda = c(\varepsilon) S(\varepsilon)^{\frac{1}{\alpha}} S(q)^{-\frac{1}{\alpha}}$ (any $\varepsilon > 0$), in effect replacing condition 1 in Proposition 2, we obtain that condition exactly as stated. We thus can replace it with such an assumption. The second part of the proposition shows that when this is the case we are actually done because the production function in the capital producing sector must be Cobb-Douglas. This crucial result closes the loop, implying that our generalized model economy can be turned into a two sector growth model featuring balanced growth and Cobb-Douglas production function in both sectors.

Definition 2. The production function in capital producing sector is Cobb-Douglas when on any interval $q \in \mathcal{I}_{ab} = [a, b] \subset \mathcal{Q}$, $a < b$, there exist $\hat{k}$ and $\hat{l}$ such that $k^e(q) = \mathbb{E}[k(q) | q \in \mathcal{I}_{ab}]$ can be represented by the cost minimization of the form $k^e \propto \min_{\hat{k}>, \hat{l}>0} r\hat{k} + w\hat{l}$ subject to $\hat{k}^{\alpha'} \hat{l}^{1-\alpha'} = 1$, where $0 < \alpha' < 1$.

Proposition 4. *Part I (necessity): Suppose the production function in capital producing sector is Cobb-Douglas in the sense of the definition above and* $\mu_q^\Lambda := c(\varepsilon) CS(\varepsilon)^{\frac{1}{\alpha}} S(q)^{-\frac{1}{\alpha}}$ *for some* $0 <$

[33] In the presence of any asymmetry between sectors, the cost $\kappa(w, r)$ will pertain to the cost in the capital producing sector q and it will not disappear as it does here.

$\alpha < 1$, then $k(q) = c(\varepsilon) CS(\varepsilon)^{\frac{1}{\alpha}} S(q)^{-\frac{1}{\alpha}} \kappa(w,r)$ almost everywhere, where $\hat{C}\kappa(w,r)$ is unit cost of production for some constant $\hat{C} > 0$. Part II (sufficiency): The aggregate production function in the capital producing sector is Cobb-Douglas when $k(q)$ is as stated in Part I.

The above result hides another win for Pareto distribution. Note that $\mu_q^\Lambda = CS(q)^{-\frac{1}{\alpha}}$ under Pareto probability density takes a simple form. For example, with $\zeta = 1$ (Zipf's law) this equation implies a constant elasticity relationship between complexity and the measure of capital-defining set equal to α; that is, $d\log q/d\log\mu_q^\Lambda = \alpha$. This is a simple relationship and simplicity is needed here to ensure that we can hope for endogenizing the distribution g induced on the space of the measures of capital-defining sets. In this case, it is simple: Pareto distributed complexity implies Pareto distributed measures of capital-defining tasks with an adjusted exponent. Pareto distribution is known to have appealing micro-foundations, as we discuss in the last section.

5 Automation breakthroughs and the labor share

Here we use the theory from the previous section to identify a form of technical progress that, if invented, lowers the labor share and yet preserves the Cobb-Douglas production function in the course of growth through automation.

The formal description is in Definition 3. We refer to this type of progress as *the universal automation-augmenting technical breakthrough,* or simply automation breakthrough. The formal definition is not straightforward, but the idea behind it is simple. This technical progress brings a new technology that, at a fixed cost of completing a small set of tasks $\mathcal{F}$, permanently "compresses" tasks within any unit capital production process by scaling the measure of sets. Formally, the key here is the assumption that for any subset $\mathcal{A} \in \Lambda_q$ this technology yields a new set $\mathcal{A}'$ (also measurable) such that the measure of the new set changes from $\tilde{\mu}_q(\mathcal{A})$ to $\tilde{\mu}_q(\mathcal{A}') = f(n)\tilde{\mu}_q(\mathcal{A})$, where $n \in \mathbb{R}_{++}$ is a variable input that goes into this transformation (a choice variable). Assumption 2 specifies the functional form for f to ensure the production function remains Cobb-Douglas. Throughout, we use the setup from our leading example to explain the effect of this technology.So the starting point is $k(q) = Z^{-1}q^\theta$ and the Pareto measure of complexity.

Definition 3. *A universal automation-augmenting technical breakthrough (in short, automation breakthrough) is a measurable set of tasks $\mathcal{F} \subset \mathcal{Q}$ with a positive measure and a Lipschitz map*[34]

[34] *A function satisfies Lipschitz condition on set D then there exists a constant $C > 0$ such that for all $x, y \in D$ $|M_n(x) - M_n(y)| \leq C|x - y|$. Lipschitz condition preserves measurability of task sets.*

$F(q,n) : \mathcal{Q} \times \mathbb{R}_{++} \rightarrow Q$ such that: 1) for all n, the function $F(\cdot, n)$ satisfies Lipschitz condition on capital defining sets $\Lambda(q)$; 2) completing tasks $\mathcal{F}$ $n \in \mathbb{R}_{++}$ times transforms the task set to produce capital of type q according to the map $F(\cdot, n)$ with the property that for any Borel set $\mathcal{A} \subseteq \Lambda(q)$, $\tilde{\mu}_q(F(\mathcal{A}, n)) = \tilde{\mu}_q(\mathcal{A}) f(n)$, where $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is some strictly decreasing function towards zero.

Assumption 2. *F is such that $f(n) = Cn^{-\beta}$, where $0 < \beta < \alpha^{-1} - 1$ and $C > 0$ is a constant.*

As shown in Proposition 5, if a breakthrough of this sort occurs, it results in a decline in the labor share by $\beta\alpha$, but the Cobb-Douglas relation between aggregate inputs is preserved after a sufficiently long spell of growth through automation. The way we think about this technology is that it is applied once at a fixed cost and it transforms the space of tasks once and for all for everybody. The firm that does the transformation is zero profit firm that has access to this technology, and so the transformation is done without a markup because multiple firms complete to make that technology available. Once it is available all firms can use it. We leave it for the feature research to extend this model so that firms have market power and actually invent this technology at a cost. Here, note, the technology arrives exogenously and we can do it under zero profits. Let us now illustrate how it works by proving the above result (the more technical bits are relegated to Appendix A as noted).

Proposition 5. *Consider the economy as in our leading example and suppose the labor share is $LS_0 = 1 - \frac{\zeta}{\theta}$. Then, with growth in automation on the balanced growth path, the post-breakthrough aggregate production function converges to a new Cobb-Douglas production function with $LS_1 = LS_0 - \frac{\zeta}{\theta}\beta$, with the initial departure from the Cobb-Douglas production function monotonically declining over time.*

Consider the capital producing process at some complexity level q, and note that a capital producing plant will choose n to minimize the cost of producing a unit of capital. Assuming it chooses to apply the technology, we have $k(q) = \min_{n \geq 0} BA_q f(n) + bn$, where we assume $A_q = Cq^{\theta}$ by Assumption (2) and by Proposition 4 we know $k \propto q^{\theta}$. The constant $B > 0$ captures the resource cost of completing tasks associated with capital of type q—and it is a constant because the composition of tasks is unchanged after the application of the breakthrough technology (the measure function is only re-scaled and the distribution of complexity within each set remains the same). Based on the previous section, we know we can write it this way, albeit B will generally depend on factor prices and there will be an equilibrium feedback here via factor prices. The constant $b > 0$ does the same for task set $\mathcal{F}$ defining the application of the technology. Whatever factor prices were determining B at the moment of the first application of this technology, that B determined what k is. It is possible that if factor prices fall significantly thereafter, the technology may be reapplied, but it must be done from scratch, in which case the fixed cost of reducing it is preventing it from happening for moderate price changes. Again, we leave these types of extensions to future work, while acknowledging that more work is required here to flesh out the effect of this technology over time.

The optimality condition is $n^* = \left(\frac{BC}{b}\mu_q^{\Lambda}\right)^{\frac{1}{1+\beta}}$, which is both necessary and sufficient assuming it is beneficial to employ this technology at a point in time ($k(q)$ goes down relative to what it was). Plugging it back into the objective function we have

$$k(q) = \frac{b}{\beta}\left(\frac{BC}{b}\mu_q^{\Lambda}\right)^{\frac{1}{1+\beta}}\left(\left(\frac{BC}{b}\right)^{\frac{1}{1+\beta}} + 1\right). \tag{21}$$

Under the assumptions made in the leading example, the functional form for capital requirement is now of the form $k(q) \propto q^{\theta\frac{1}{1+\beta}}$, with all the constants of proportionality "~" soaked up in the TFP (which also rises). If this capital requirement function applied globally to all q's, it would have given rise to a Cobb-Douglas production function with a lower labor share. It is clear from the leading example that the labor share would have declined from $LS_0 = 1 - \frac{\zeta}{\theta}$ to $LS_1 = LS_0 - \beta\frac{\zeta}{\theta}$.

However, it may not be optimal to apply this technology. The above problem implies that there exists a cutoff $q_{\min} > 0$ such that the technology is applied only to capital production above this threshold. The lower b/B is, the lower the threshold, but it is never zero. Note that the log benefit from applying this technology is proportional to the mass of the production task set A_q, and so it will never be applied to tasks with a very low μ_q^{Λ}, but if it is optimal to apply to a given q it is applied to all q above that level. Therefore, the new capital requirement applies above some $q_{\min}$, and below we still have $k(q) \propto q^{\theta}$. Importantly, the two pieces connect at q_{min} because the choice of this threshold is optimal (there is no discontinuity). The fact that the new capital requirement does not apply to the lower portion of the domain introduces an isoquant error with respect to the Cobb-Douglas isoquant. The rest of the proof shows that this error—just like the one introduced by $q_0 > 0$ in the leading example in Section 2—vanishes with growth in automation. This part is technical and has been relegated to Appendix A. Intuitively, as q^* rises with growth in automation, the lower portion of the $k(q)$ function becomes (relatively) unimportant.

Automation breakthroughs in our model can be stacked and occur repeatedly, with each subsequent breakthrough reducing the labor share from the previous level and leading to another Cobb-Douglas production function after convergence. As mentioned, it would be natural to think about R&D process that invents such technologies, but this goes beyond the scope of a single paper.

Is the second machine age an automation breakthrough? A discussion

Do the properties of an automation breakthrough in our model accord well with the anecdotal description of the kind of transformative technologies that are suspect of reducing the labor share, such as AI or ICT? The impact of the new generation of these technologies lead Brynjolfsson and McAfee (2014) to refer to the modern automation era as "the second machine age," and distinguish it from the previous phases of the industrial revolution labeled jointly as "the first machine age". A case can be made that some of these technologies, as discussed by these authors, do have the characteristics of a breakthrough technology in our model. Let us develop this point further.

Note that the main characteristics of a breakthrough technology in our model are: 1) a fixed cost of application to production of capital, 2) universal applicability, 3) task compression, and 4) scalability. The first property means that an application of the breakthrough technology entails a fixed cost, which cannot be too large so that the technology is broadly applied. The second property means that the breakthrough technology *can* be applied to most types of capital (tasks). The two are crucial to change $k(q)$ schedule almost for most values of $q \geq q_{\min}$, with $q_{\min}$ not too large. The third property is that it "compresses" the task space underlying the production process of capital by eliminating a fraction of tasks. This key feature flattens capital requirement function $k(q)$ and leads to lower labor share. The last property means that the technology is scalable; that is, a bigger effect can be achieved at a higher fixed cost—potentially subject to decreasing returns to scale ($\beta < 1$). This last feature is only to preserve the shape of $k(q)$ so that it remains of constant elasticity with respect to q.

Let us now provide two illustrative examples that are suggestive of these features. These examples are in no way exhaustive. They are here to motivate our model and stimulate further discussion based on concrete theory.

Consider first the growing use of advanced industrial robots in a factory setting, which is a well-documented and accelerating phenomenon (Graetz and Michaels, 2018). Simplifying a bit, a computer chip augmented with AI can give the same metal arm the ability to perform a much a broader range of tasks by processing sensory information on par with humans. Such robots can be used in unstructured environments, work alongside humans, and perform a broader range of tasks. The technology is universal because all kinds of tasks require processing sensory information. The cost of an AI-augmented robot is roughly the same and it boils down to the application of AI software, and an appropriate sensor suite. While the cost of developing IT for industrial robots is substantial, over time it has the potential to become negligible due to massive increasing returns to scale. Conditions 1 and 2 are thus fulfilled.

In terms of task compression, the key observation is that a single AI-augmented robot can be employed to a whole "class" of tasks rather than something specific, which effectively means the cost of providing capital declines per affected task. The payoff is also bigger the more diverse the environment is, which previously corresponded to its complexity presented to robots. This means that condition 3 is present too. For example, in a car factory setting, an AI-augmented industrial robot system may perform any type of weld instead of a particular type of weld, turning all welds to a "class of one." For a human, a weld is a weld, but for robots of the past there were hundreds of different welds demanding specific accommodations. IThis means that the costs of providing "welding capital" fell because one robot can substitute for a number of custom robots from the past. The technology is also scalable. More complex systems can deliver more at a higher cost, and feature 4 is also present in some form.

ICT has a similar flavor, although the link is less direct. The key aspect, as we see it, is that ICT moves tasks to the cyberspace and this way compresses tasks involved in the production of automation capital. The synergy with other technologies is important regarding its big picture role. For example, an Uber driving service app navigates the car, matches drivers with customers, provides security by tracking the car, automates payments, informs the passenger about the position of the car, collects ratings, manages driver's time, sets instantaneous demand-dependent prices, and so on and so fourth, all using a single platform of a phone, software and wireless connectivity. In the past the underlying set of tasks to produce a piece of capital to "assist driving a taxi passenger" would require many separate pieces of machinery, and here a few lines of codes can do it all for the entire globe.

It may be too early to tell whether we are experiencing the next phase of the industrial revolution, but it is clear that a new quality of capital has emerged and that is well-suited suitable for complex and unstructured environments. Of course, sophisticated machines existed in the past. For example, a power loom is one impressive machine. But such examples so far could be found in structured environments; in fact, the biggest revolution in automation came from turning unstructured environments to structured environments where machines could be employed (e.g., assembly line). At the very least, capital during the second machine age has the potential to spread further by breaking away from this constraint.

6 Endogenous distribution of tasks

Keynes once quipped that the constancy of labor share in the data is "a bit of a miracle." It is the premise of our analysis that nature does not rely on miracles, only its basic laws and principles may imply ones. Our analysis leaves our one important "miracle" unexplained: Pareto distributed task complexity or, equivalently, the measure of capital-defining sets.

It is well known that Pareto distribution arises in nature. From this point of view, then, it is not much of a miracle, and we just may not yet understand how it arises in production of capital goods. We leave this largely for future research, but finish by discussing several concrete examples of R&D processes to convince the reader that the known ways of obtaining Pareto distribution are compatible with our analysis. The discussion draws on Newman (2004) and especially Gabaix (2009).

Random growth. Our first example employs random growth—one of the simplest mechanism to obtain power law dynamics. To adopt this approach, suppose that task complexity on average declines at a rate $\gamma < 1$ per unit of time. That is, $\bar{q}_{t+1} = \gamma \bar{q}_t$, where $\bar{q}_t$ is the mean complexity across all tasks. Furthermore, assume the distribution of the decline is uneven across individual tasks because innovations affect individual tasks differently. Some tasks may not decline at all in a given period, in which case their relative complexity rises by $\eta = \frac{1}{\gamma}$ relative to the average, while other may decline by more than the average. The important assumption is that this process is i.i.d. across tasks. The rest follows from the discussion of random growth model in Gabaix (2009).

Inverse function. The fact that capital requirement is an inverse of productivity of capital can be used to trivially obtain tail power law from the basic fact of taking the inverse (Sornette, 2002). Let $y = x^{\frac{1}{\zeta - 1}}, \zeta - 1 > 0$ and suppose x is distributed according to some pdf $p_x(x)$ such that $p(x) \to C > 0$ as $x \to 0$. Then, the tail distribution of y follows a power law with exponent ζ. Of course, to apply this result it must be that the economy operates far into the tail of the distribution. The fact that our model requires warped distribution adds extra flexibility of what is allowed away from the tail.[35]

Yule process. Another appealing idea explaining the prevalence of Pareto distribution is the one underlying Yule's "speciation" process. The key to this approach is the observation that a variable that grows exponentially and is stopped after an exponentially distributed time is Pareto distributed

[35] The result follows from the change of variables formula.

at the stopping time.[36] We relegated this extension to the Online Appendix C because it is fairly involved. The framework we develop there is based on information theory and it is establishes a bridge between the current model and the combinatorial growth literature (Weitzman, 1998; Jones, 2021).

7 Conclusions

This paper contributes an analytically tractable task-based theory of the Cobb-Douglas production function that serves as a building block for many economic models and is the foundation for balanced growth when technological progress is not only labor augmenting. Our microfounded theory of production explains the coexistence of growth through automation and balanced growth in the neoclassical framework in a descriptively realistic manner, while preserving its desirable properties in macro applications. We have characterized the endogenous link between our theory's fundamentals and the labor share, including the characterization of labor share-reducing technical changes. We left the question of how the technology frontier of the economy evolves for future research, but we have demonstrated that the known ways of obtaining a Pareto distribution can be integrated into our theory. In terms of substantive insights, we have shown that a technical innovation that makes capital better in dealing with marginal increase in task complexity can lead to a permanently lower level of the labor share on the balanced growth path. We have proposed an explicit analytic formulation of such breakthrough technologies and discussed how their characteristics relate to AI and ICT—the key transformative technologies that are widely believed to have taken automation to the next level in the modern era. We characterized the endogenous forces that stabilize the labor share in the course of growth through automation.

References

Acemoglu, D. (2003): "Labor- and Capital-Augmenting Technical Change," *Journal of the European Economic Association*, 1, 1–37.

[36] The key mathematical property is that an exponential of an exponentially distributed random variable is Pareto distributed, as the following calculation shows ($X \sim$exponential, $Y = \exp(X)$):

$$Pr\left(Y < y\right) = Pr\left(\exp\left(X\right) \leq y\right) = Pr\left(X \leq \log\left(y\right)\right) = 1 - x^{-\lambda}.$$

Acemoglu, D. and D. Autor (2011): "Skills, Tasks and Technologies: Implications for Employment and Earnings," in *Handbook of labor economics*, Elsevier, vol. 4, 1043–1171.

Acemoglu, D. and P. Restrepo (2018): "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment," *American Economic Review*, 108, 1488–1542.

——— (2020): "Robots and Jobs: Evidence from US Labor Markets," *Journal of Political Economy*, 128, 2188–2244.

Armenter, R. (2015): "A Bit of a Miracle No More: The Decline of the Labor Share," *Federal Reserve Bank of Philadelphia Business Review*, 98.

Arntz, M., T. Gregory, and U. Zierahn (2016): "The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis," *OECD Social, Employment and Migration Working Papers*.

Atkeson, A. and P. J. Kehoe (2007): "Modeling the Transition to a New Economy: Lessons from Two Technological Revolutions," *American Economic Review*, 97.

Autor, D. and A. Salomons (2018): "Is Automation Labor-Displacing? Productivity Growth, Employment, and the Labor Share," *Brookings Papers on Economic Activity*, BPEA Conference Drafts.

Autor, D. H., F. Levy, and R. J. Murnane (2003): "The Skill Content of Recent Technological Change: An Empirical Exploration," *The Quarterly journal of economics*, 118, 1279–1333.

Azariadis, C. and J. Stachurski (2005): "Poverty Traps," in *Handbook of Economic Growth*, ed. by P. Aghion and S. Durlauf, Amsterdam: Elsevier, vol. 1A.

Billingsley, P. (1995): *Probability and Measure*, John Wiley and Sons, Inc., 3 ed.

Brynjolfsson, E. and A. McAfee (2014): *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, WW Norton & Company.

Dao, M. C., M. M. Das, Z. Koczan, and W. Lian (2017): *Why is Labor Receiving a Smaller Share of Global Income? Theory and Empirical Evidence*, International Monetary Fund.

David, H. and D. Dorn (2013): "The Growth of Low-skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, 103, 1553–97.

Dvorkin, M. A. and A. Monge-Naranjo (2019): "Occupation Mobility, Human Capital and the Aggregate Consequences of Task-Biased Innovations," *FRB St. Louis Working Paper*.

Elsby, M. W. L., B. Hobijn, and A. S. Sahin (2013): "The Decline of the U.S. Labor Share," *Brookings Papers on Economic Activity*, Fall.

Evenson, R. and Y. Kislev (1976): "A Stochastic Model of Applied Research," *Journal of Political Economy*, 84.

Ford, M. (2009): *The Lights in the Tunnel: Automation, Accelerating Technology and the Economy of the Future*, CreateSpace Independent Publishing Platform.

Frey, C. B. (2020): *The Technology Trap: Capital, Labor, and Power in the Age of Automation*, Princeton University Press.

Gabaix, X. (2009): "Power Laws in Economics and Finance," *Annual Review of Economics*, 1.

Ghiglino, C. (2012): "Random Walk to Innovation: Why Productivity Follows a Power Law," *Journal of Economic Theory*, 147.

Gomme, P. and P. Rupert (2004): "Measuring Labor's Share of Income," *Federal Reserve Bank of Cleveland, Policy Discussion Paper*, 04-07.

Goos, M. and A. Manning (2007): "Lousy and Lovely Jobs: The Rising Polarization of Work in Britain," *Review of Economics and Statistics*, 89, 118–133.

Graetz, G. and G. Michaels (2018): "Robots at Work," *Review of Economics and Statistics*.

Greenwood, J., Z. Hercowitz, and P. Krusell (1997): "Long-Run Implications of Investment-Specific Technological Change," *The American Economic Review*, 87.

Grossman, G. and E. Helpman (1991): *Innovation and Growth in the Global Economy*, Cambridge: MIT Press.

Grossman, G. M., E. Helpman, E. Oberfield, and T. Sampson (2017): "Balanced Growth Despite Uzawa," *American Economic Review*, 107.

Growiec, J. (2013): "A Microfoundation for Normalized Ces Production Functions with Factor-Augmenting Technical Change," *Journal of Economic Dynamics and Control*, 37.

Gutierrez, G. and S. Piton (2020): "Revisiting the Global Decline of the (Non-housing) Labor Share," *American Economic Review: Insights*, 2.

Hansen, G. D. and E. C. Prescott (2002): "Malthus to Solow," *The American Economic Review*, 92.

Houthakker, H. S. (1955): "The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis," *Review of Economic Studies*, 23.

Jones, C. I. (2005): "The Shape of Production Functions and the Direction of Technical Change," *Quarterly Journal of Economics*, 120.

——— (2021): "Recipes and Economic Growth: A Combinatorial March Down an Exponential Tail," *unpublished manuscript*.

Jones, C. I. and P. M. Romer (2010): "The New Kaldor Facts: Ideas, Institutions, Population, and Human Capital," *American Economic Journal: Macroeconomics*, 2.

Jones, C. I. and D. Scrimgeour (2008): "A New Proof of Uzawa's Steady-State Growth Theorem," *The Review of Economics and Statistics*, 90.

Kaldor, N. (1961): "Capital Accumulation and Economic Growth," in *The Theory of Capital*, ed. by F. Lutz and D. Hague, St. Martin's Press.

Karabarbounis, L. and B. Neiman (2013): "The Global Decline of the Labor Share*," *The Quarterly Journal of Economics*, 129, 61–103.

Koh, D., R. SantaeulÃ lia-Llopis, and Z. Yu (2020): "Labor Share Decline and Intellectual Property Products Capital," *Econometrica*, 88.

Kortum, S. S. (1997): "Research, Patenting, and Technological Change," *Econometrica*, 65.

Leon-Ledesma, M. A. and M. Satchi (2019): "Appropriate Technology and Balanced Growth," *Review of Economic Studies*, 86.

Manyika, J., J. Mischke, J. Bughin, J. Woetzel, M. Krishnan, and S. Cudre (2019): "A New Look at the Declining Labor Share of Income in the United States," *McKinsey Global Institute*.

Michaels, G., A. Natraj, and J. Van Reenen (2014): "Has ICT Polarized Skill Demand? Evidence from Eleven Countries over Twenty-Five Years," *Review of Economics and Statistics*, 96, 60–77.

Muro, M., R. Maxim, and J. Whiton (2019): "Automation and Artificial Intelligence: How Machines Are Affecting People and Places," *Metropolitan Policy Program at Brookings*.

Newman, M. (2004): "Power Laws, Pareto Distributions and Zipf's Law," *Contemporary Physics*, 46.

Perla, J. and C. Tonetti (2014): "Equilibrium Imitation and Growth," *Journal of Political Economy*, 122.

Romer, P. (1986): "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, 94.

Schlicht, E. (2006): "A Variant of Uzawa's Theorem," *Economics Bulletin*, 5.

Sornette, D. (2002): "Mechanism for Powerlaws without Self-Organization," *International Journal of Modern Physics C*, 13.

Uzawa, H. (1961): "Neutral Inventions and the Stability of Growth Equilibrium," *Review of Economic Studies*, 28.

Weitzman, M. L. (1998): "Recombinant growth," *Quarterly Journal of Economics*, 113.

Wheeden, R. L. and A. Zygmund (1977): *Measure and Integral: An Introduction to Real Analysis*, Marcel Dekker, Inc.

Appendix A: Omitted proofs

This main appendix contains omitted proofs of the main results in text. (The proof of Lemma 1, and Corollary (2) are in the Online Appendix B.)

Proof of Proposition 2: Let $q_0 := \inf Q$, as defined in the setup. I. Necessity: Let K be total input of capital used by the plant, let L be total input of labor, and let Y be the output. If the relationship between Y, K and L is Cobb-Douglas with exponent α ($Y = AK^{\alpha}L^{1-\alpha}$, where $0 < \alpha < 1, A > 0$), the relationship between factor intensities is

$$\left(\frac{K}{Y}\right)^{\alpha}\left(\frac{L}{Y}\right)^{1-\alpha} \equiv \frac{1}{A}. \tag{22}$$

This equation is an identity, as indicated by "$\equiv$;" that is, it must hold for all $K/Y > 0, L/Y > 0$, given some $A > 0$. In our model this identity is required hold for inputs which are cost minimizing, since we are solving for a production function implied by cost minimization. As a result, the above identity is a restriction on optimal factor intensities $\frac{K}{Y}\left(\frac{w}{r}\right)$ and $\frac{L}{Y}\left(\frac{w}{r}\right)$ implied by cost minimization. The identity must hold for all w/r.

Plugging in from (4) and (5), equation (22) implies we need

$$\left(\int_{[q_0,q^*]} k\left(q\right)gdv\right)^{\alpha}\left(S\left(q^*\right)\right)^{1-\alpha}\equiv\frac{1}{A}.$$

Since we require this to be an identity (all values of q^*), we can take a Lebesgue derivative with respect to q^* to obtain

$$\alpha\left(\int_{[q_0,q^*]} k\left(q\right)gdv\right)^{\alpha-1} k\left(q^*\right)g\left(q^*\right)\left(S\left(q^*\right)\right)^{1-\alpha}$$
$$-\left(1-\alpha\right)\left(\int_{[q_0,q^*]} k\left(q\right)gdv\right)^{\alpha}\left(S\left(q^*\right)\right)^{-\alpha}g\left(q^*\right)\equiv 0,$$

which simplifies to

$$\int_{[q_0,q^*]} k\left(q\right)gdv\equiv\frac{\alpha}{1-\alpha}k\left(q^*\right)\left(S\left(q^*\right)\right). \tag{23}$$

Since the above relationship needs to hold up to an arbitrary constant $Z>0$, the differential equation is both necessary and sufficient.

Next, we differentiate again the above expression with respect to q^*, while noting for later that the solution of the obtained differential equation this time will give us a necessary condition up to a constant that will need to take a particular value (we will return to this). Differentiation gives

$$k\left(q\right)g\left(q\right)=\frac{\alpha}{1-\alpha}k'\left(q\right)S\left(q\right)-\frac{\alpha}{1-\alpha}k\left(q\right)g\left(q\right), \tag{24}$$

where the asteriks over q is no longer needed since we require this to hold for any q^* (any factor prices—a point we return to at the end). Simplifying the above expression, we obtain condition 2, $\alpha\frac{k'(q)}{k(q)}=h(q):=\frac{g(q)}{S(q)}$, which is well-defined by condition 1.

We now return to sufficiency.

II. Sufficiency: The disconnect between necessity and sufficiency is brought about by: 1) integration constant implied by differentiation of (23), and 2) the fact that we implicitly assumed an interior and unique solution underlying $q_k^*\left(\frac{w}{r}\right)$ in (23). We consider them one by one.

II.1. Solution of the differential equation (24) implies (23) up to a constant and we need this constant to be zero. This means that we are seeking a particular solution with C in equation (25) such that (24) holds. What are they requirements on function k to ensure that constant is zero? We will use $q^* \to q_0$ to identify it. It is clear from (23) that $k(q^*) \to_{q^* \to q_0} 0$ this is the case, since integrability of k on domain $\mathcal{Q}$ implies $\int_{q_0}^{q^*} k(q^*)g(q^*) \to_{q \to q_0} 0$. Accordingly, both the right-hand side in (24) ($\lim_{q^* \to q_0} \frac{\alpha}{1-\alpha} k\left(q^*\right)\left(1 - G\left(q^*\right)\right) = 0$) and the left-hand side ($\lim_{q^* \to q_0} \int_{q_0}^{q^*} k\left(q\right) g\left(q\right) = 0$) vanishes in the limit. (The last property follows from the mean value theorem and existence of the integral.) Concluding, we need $k\left(q\right) \to_{q \to q_0} 0$, as stated, or equivalently $C \to 0$.

II.2. Deriving (1), we assumed an interior and unique solution on the entire domain of w/r, which can take any value on the set $(0, \infty)$. This requires that on the entire domain we must be able to solve $k(q^*) = w/r$. Using equation (25) (and the differential equation preceding it), the function $k(q)$ must be invertible, and so it must be strictly increasing. Consequently, g must have full support on $\mathcal{Q}$ by (14). For the solution to exist for low values of w/r, we need $k\left(q_0\right) \to_{q \to q_0} 0$, which we imposed in II.1. We also need $k(q)$ to be unbounded since w/r may be unbounded, but this is automatically ensured by (15) and the fact that $1 - G\left(q\right)$ is converging to 0 as $q \to q_0$. Concluding, we only need to add that g has full support, and the rest is taken care of by the fact that then $S\left(q\right) \to_{q \to q_0} 0$. (It is possible to have $S = \infty$ for all q, but this is under the trivial measure and it is not permissible as stated.) Q.E.D.

Proof of Lemma (2) The unique solution to ODE in condition 2 exists and is of the form

$$k\left(q\right) = c\left(\varepsilon\right) C \exp\left(\alpha^{-1} H_\varepsilon\left(q\right)\right), \tag{25}$$

where $H_\varepsilon\left(q\right) := \int_\varepsilon^q h\left(u\right) du$, and where $c\left(\varepsilon\right)$ is a constant implied by the solution of ODE that will need to take a particular value for each $\varepsilon > 0$ up to a positive scaling constant C to ensure this formula implies the same value for any ε for a fixed q (the fact it is arbitrary implies it is a function of ε, no more). Since it is possible that $S\left(q\right) \to \infty$ at $q = q_0$ (see exact balance growth in our leading example with $q_0 = 0$), we must be careful at the lower bound of the distribution, and hence we bounded it away by $\varepsilon > 0$. The relationship between the cumulative hazard function and the distribution G is

$$H_{\varepsilon}(q) = \int_{\varepsilon}^{q} \frac{g(u)}{S(u)} du = \int_{\varepsilon}^{q} \frac{1}{S(u)} \underbrace{\left(-\frac{d}{du}(S(u))\right)}_{\equiv g(u)} du = \ln\left(\frac{S(\varepsilon)}{S(q)}\right).$$

Accordingly, $k(q) = c(\varepsilon) CS(\varepsilon)^{\frac{1}{\alpha}} S(q)^{-\frac{1}{\alpha}}$, for all $q \geq \varepsilon + q_0$ and given any $\varepsilon > 0$, where $C > 0$ is some positive constant.

Proof of Corollary 1: The conclusion follows directly from the lemma above after taking $c(\varepsilon) \to 0$, which is possible here because S is bounded by 1 when g is probability density.

Proof of Proposition 3: Let $S(q;p) := 1 - G(q;p)$ be the counter cumulative distribution function (survival function) associated with G on some domain $[q_0, \infty)$,where $q_0 \geq 0$. Let G be a parameterized family of functions by some real number p. For example, in the case of Pareto distribution, we have $G(q) = 1 - (q_0/q)^{\zeta}$, and p could be either ζ or q_0. We will first prove a technical result about distribution functions in general and then proceed with the proof of this result.

Lemma 3. *Let $G(q;p)$ be a parameterized family ($p \in \mathbb{R}$) of distribution functions on domain $[q_0, \infty)$, where p can be $q_0 \geq 0$, with a well-defined probability density $g(q;p)$. Suppose that: 1) there exists a function $v(p_n) : \mathbb{N} \to \mathbb{R}$ defined on an infinite sequence $\{p_n\}_{n\in\mathbb{N}}$ such that $p_n \to_{n\to\infty} p_\infty \in \mathbb{R}$,* with $v(p_n) \to_{n\to\infty} 0^+$, *and the counter cumulative distribution function (survival function) has the property that for all q on the domain*

$$S(q;p_n) := 1 - G(q;p_n) = v(p_n)\left(1 - \hat{G}(q)\right),$$

for some function $\hat{G}(q)$ independent of p_n (not necessarily a cumulative distribution function), and 2) the hazard function

$$h(q) := \frac{g(q;p_n)}{1 - G(q;p_n)}$$

is independent of p_n for all q as indicated. Then, p is q_0 (p pertains to the lower bound of the support of the distribution g).

Proof. (Lemma 3) Suppose, by the way of contradiction, that: i) The factorization of S applies and so there exists some function $v(p_n)$ such that, given the identity

$$S(q;p_n) \equiv v(p_n)\underbrace{\left(\frac{1}{v(p_n)} - \frac{G(q;p_n)}{v(p_n)}\right)}_{\equiv B(q)}, \tag{26}$$

the term identified as B is independent of the sequence p_n for all q. Moreover, ii) the hazard rate $h(q) = \frac{g(q;p_n)}{v(p_n)B(q)}$ is independent of p_n. However, by contradiction, iii) p is not q_0.

Note that the fact that h and B do not depend on p_n implies we must be able to write $g(q;p) = v(p)\,\hat{g}(q)$, for some function $\hat{g}$ independent of p_n (this immediately follows from the formula for the hazard function above and requirement ii). But, since p is not q_0 by iii, we can write

$$G(q^*;p) \equiv \int_{q_0}^{q^*} g dq = v(p)\int_{q_0}^{q^*} \hat{g}(q)\,dq = v(p)\,\hat{G}(q^*),$$

for some function $\hat{G}$ that is also independent of p_n by the hypothesis. This is a contradiction because B defined in (26) ends up depending on p, since $B(q^*) = \frac{1}{v(p_n)} - \hat{G}(q^*)$. Hence, p must be q_0. Q.E.D. □

This lemma is key because the constant C in equation (15) must converge to zero by condition 3 (see Corollary 1), and yet k cannot be zero everywhere because this leads to a contradiction that we are in $\mathcal{T}$. There is one way to avoid this (see our leading example). For k to be a nonzero well-defined function, there must exist a parameter p and a sequence $\{p_n\}_{n\in\mathbb{N}}$ on that parameter space such that there is a function $v(p_n)$ with the property $v(p_n) \to 0^+$ as $p_n \to_{n\to\infty} p_\infty \in \mathbb{R}$ and $S(q;p_n) = v(p_n)\,\hat{S}(q)$, for some continuous and decreasing function $\hat{S}(q)$ on $\mathbb{R}_{++}$ that does not depend on p_n, and actually such that $\hat{S}(q) \to_{q\to 0} \infty$ (we do not need it here but it will come up at the end). By Corollary 1, we then can have $C_n \to 0^+$ and yet ensure $k(q) = C_n v(p_n)^{-\frac{1}{a}}\,\hat{S}(q)^{-\frac{1}{\alpha}}$ is independent of p_n and that it is not zero everywhere. Note that, if $C_n v(p_n)^{-\frac{1}{a}}$ can be made into a constant, then $\hat{S}(q) \to_{q\to 0} \infty$ suffices to ensure this result. Without loss we can assume that $v(p_n)$ is positive-valued. Now, it is important to stress that this must be true except for finitely many terms of the sequence only, in which case anything is possible. But it can not be false on any sub-sequence (countable number of terms).

The lemma above shows that the only parameter that can be factored out this way is the lower bound of the support of the distribution, that is p must be q_0 (or some function of it that can be absorbed into v). This restricts what G can be. By the fact that $G(q,p) = 1 - v(p_n)\hat{S}(q)$,where $\hat{S}$ is positive-valued function independent of p_n, strictly decreasing, and continuous in q—hence also differentiable in q (almost everywhere). Then, since p is q_0 by Lemma 3, by the fundamental theorem of calculus, we must have

$$G(q^*; p_n) := -\int_{p_n}^{q^*} v(p_n)\hat{S}_q(q)\,dq \equiv v(p_n)\hat{S}(p_n) - v(p_n)\hat{S}(q^*),$$

and since $S(q^*; p_n) \equiv v(p_n)\hat{S}(q^*)$, and S is a counter cumulative distribution with $G(\inf \mathcal{Q}) = 0$, we need

$$v(p_n)\hat{S}(p_n) = 1, \tag{27}$$

which implies $G(q^*, p_n) = 1 - \frac{\hat{S}(q^*)}{\hat{S}(p_n)}$. Since $v(p_n) \to_{n\to\infty} 0^+$, by (27), we indeed have $\lim_{q\to p_\infty} \hat{S}(q) = \infty$, as needed. Since G must be a distribution function we must have $\lim_{q^*\to\infty} \hat{S}(q^*)^{-1} = \infty$. We have now established the existence of a strictly increasing function $f(q) := \hat{S}(q)^{-\frac{1}{\zeta}}$ $(\zeta > 0)$ that meets the requirements of the lemma and the functional form for G is as stated. We have established the existence of a strictly increasing function $f(q) := \hat{S}(q)^{-1}$ that meets the requirements of the lemma and the functional form for G is as stated. Recall that this result holds except for finitely many terms of the sequence.[37] Q.E.D.

Proof of Proposition 4: Part I. Necessity: Let $q_0 := \inf \mathcal{Q} \geq 0$. Consider first the equation (18):

$$k(q) := \mu_q^\Lambda \left(\int_{q_0}^{q^*} rk(q')\,\tilde{g}^0_{i(q)}dv + w\tilde{S}^0_{i(q)}\left(q^*\left(\frac{w}{r}\right)\right)\right), \tag{28}$$

where $\tilde{g}^0_{i(q)}$ reflects Assumption 2 and asserts that there may be an i.i.d. mapping between q and the measure across tasks. Since $i(.)$ was defined as a step function, the resulting function k will be of bounded variation, and k is measurable. We can remove $i(.)$ on any line segment because it adds an i.i.d. error, no more. We first formalize this idea. Take any interval $q \in \mathcal{I}_{ab} := (a,b)$ on the complexity space. Use the standard approximation of k as a step function from below on that interval so that it converges uniformly to k on that interval.[38] The step function is constant on the flat segments of the domain. On the flat segments of the step function we can thus replace $\tilde{g}^0_{i(q)}(.)$ by

[37]Suppose not, then there exists a subsequence that converges to the desired result and it is not the one given above, which we have shown is not possible.

[38]See Appendix B.1 for a formal definition of this step function for k and note that the existence of such an approximating step function is a standard result in measure theory, particularly suitable if we convert the above integral to Lebesgue integrals.

$\tilde{g}(q') := \mathbb{E}\left[\tilde{g}^0_{i(q)}(q')\,|q \in \mathcal{I}_{ab}\right]$, as implied by Assumption 2, since these line segments integrate over an interval. After convergence of the step function, we obtain k^e that coincides with k except for a countable number of points, and we have

$$k^e := \mu_q^\Lambda \left(\int_{q_0}^{q^*} r k^e (q')\, \tilde{g} dv + w \tilde{S}\left(q^*\left(\frac{w}{r}\right)\right)\right), \tag{29}$$

From the proof of Proposition 2 observe that that Cobb-Douglas aggregation obtains in the capital sector on the interval $q \in \mathcal{I}_{ab}$ if and only if

$$\int_{q_0}^{q^*} r k^e (q)\, \tilde{g} dv \equiv \frac{\alpha}{1-\alpha} r k^e (q^*)\, \tilde{S}(q^*) \,. \tag{30}$$

To see this, let Y in the part I of that proof be the number of capital goods produce by the type q capital-producing sector, let K be total capital input that that sector is using, and let L be the total labor it uses. Multiply both sides by r. We have shown in text examples that this condition can be satisfied, and existence does not pose a challenge. This gives us the integral on the left-hand side of (29).

Recall that $q^*\left(\frac{w}{r}\right) = k^{e-1}\left(\frac{w}{r}\right)$, and so $\tilde{S}\left(q^*\left(\frac{w}{r}\right)\right) = \tilde{S}\left(k^{e-1}\left(\frac{w}{r}\right)\right)$. This condition, note, involves the inverse $k^{e-1}\left(\frac{w}{r}\right)$ that is well-defined because we have now removed the i.i.d. component and this function is strictly increasing (as we will show at the end, so for now proceed). Plugging in to (29), after substitutions using the above two equations, we obtain

$$k^e (q) = \mu_q^\Lambda \tilde{S}(q^*) \left(\frac{\alpha}{1-\alpha} r k^e \left(k^{e-1}\left(\frac{w}{r}\right)\right) + w\right)$$

hence $k^e (q) = \mu_q^\Lambda \tilde{S}(q^*) \frac{w}{1-\alpha}$ and

$$k^e (q) := A_q \kappa (w, r) \,, \tag{31}$$

where $\kappa (w, r) = \hat{C} \tilde{S}\left(k^{e-1}\left(\frac{w}{r}\right)\right) \frac{w}{1-\alpha}$ and $\hat{C} > 0$ is some constant. As noted in the leading example, this term corresponds to the unit cost production cost, here up to a constant because at this intermediate step we are not yet including the normalization by total factor productivity $(1/A)$. Although k^{e-1} is involved in this expression, this is a function of factor prices and since it is strictly increasing here, it is globally invertible. Plugging in for μ_q^Λ from the statement of the proposition obtains the result.

Part II. Sufficiency: All steps in the necessity part were "if and only if," and so the reasoning can be reversed. We thus have established sufficiency. For an explicit proof, see Online Appendix B. Q.E.D.

Proof of Proposition 5: The first part is in text. Observe that the breakthrough technology is applied to capital producing task for all q such that $\frac{b}{\beta}\left(C\frac{B}{b}\mu_q^{\Lambda}\right)^{\frac{1}{1+\beta}}\left(\left(C\frac{B}{b}\right)^{\frac{1}{1+\beta}}+1\right)\leq B\mu_q^{\Lambda}$,where we know $\mu_q^{\Lambda}\propto q^{\theta}$ because we assume the setup from the leading example. It is clear that there exists a finite cutoff q_{min} such that the technology is applied for all $q\geq q_{min}$ ($\beta>0$). We know from the leading example that $k\left(q\right)=Cq^{\theta}$ for $q\leq q_{min}$ and $k\left(q\right)=Cq^{\theta(1+\beta)}$ otherwise. Consider now the integral that defines capital intensity after the breakthrough. It is

$$\frac{K}{Y}=A^{-1}C\zeta\left(\int_{q_o}^{q_{min}}q^{\theta-\zeta-1}dq+\int_{q_{min}}^{q^*}Cq^{\theta(1+\beta)-\zeta-1}dq\right).$$

Define the error as the difference relative to counterfactual of new $k\left(q\right)=Cq^{\theta(1+\beta)}$ applying to the entire domain; that is, let $\frac{K'}{Y'}=A^{-1}C\zeta\left(\int_{q_0}^{q^*}q^{\theta(1+\beta)-\zeta-1}dq\right)$, take limit $q_0\to 0$, and define the isoquant error as

$$\varepsilon:=\frac{\frac{K}{Y}-\frac{K'}{Y'}}{\frac{K'}{Y'}}=\left(q^*\right)^{\zeta-(1+\beta)\theta}q_{min}^{\theta-\zeta}\frac{\left(\left(1+\beta\right)\left(\theta\left(1+\beta\right)-\zeta\right)-\left(\theta-\zeta\right)\right)}{\left(\theta-\zeta\right)}.$$

Growth through automation increases the cutoff q^* by definition, and hence $\varepsilon\to 0$ as $q^*\to\infty$.[39] Q.E.D.

[39]Derivation of the above expression can be found at the end of online Mathmatica notebook "Proposition_breakthrough.nb".

Online Appendix B: Supplementary results

This appendix contains supplementary results for the theory.

1. Normalization of labor requirement (Section 1)

Here we show that labor requirement set to 1 across all tasks can be though of as a normalization in a more general setup. This is related to the setup of the model in Section 1.

Suppose there is a separate capital and labor requirement function $\hat{k}(q)$ and $l(q)$. Define $k(q) := \hat{k}(q)/l(q)$. Let q be ordered so that $k(q)$ is increasing as assumed in text. Suppose these function are measurable and labor requirement is independent of (relative) capital requirement; that is, knowing k gives no information about l, and so for any q we have $\mathbb{E}(l(q)|k(q)) = \mathbb{E}l(q)$ (we use expectation operator here under any some arbitrary probability measure p on the task space).

This implies that there exists a constant $C > 0$ such that for any interval $I \subset \mathcal{Q}$ we have

$$\int_I l(q)\, d\mu = \int_I C d\mu,$$

since l is i.i.d. on the reordered set wrt k (if this is not the case it would be possible to infer l from k—that is, from q that $k(q)$ is increasing in—and we assume that it is not constant for all q). We assume $l(q)$ is bounded and hence integrable.

Let us now normalize units for both capital and labor requirements by C; that is, we define $\hat{k}(q) := \hat{k}(q)/C$ and $l(q) := l(q)/C$. Note that this does not change the definition of k only the units in which inputs are measured. Independence guarantees that we can recover capital usage from the ratio, since for any integral on arbitrary measurable set I we have

$$\int_I \hat{k}(q)\, d\mu = \int_I k(q)\, l(q)\, d\mu = \int_I k(q)\, d\mu \int_I l(q)\, d\mu = \int_I k(q)\, d\mu.$$

This completes the proof because $k(q)/l(q)$ is the relevant object in the model, which—once compared to the relative price w/r—suffices to determine which tasks are completed by capital and which are completed by labor.[40]

[40] That is, we do not lose relevant information by dropping $l(q)$. More generally, this may not suffice and depends on the optimization problem at hand. As a counterexample, suppose the firm—for whatever reason—chooses to do tasks with capital iff $l \geq 5$. In that case the ratio k/l would not be sufficient.

2. Derivation of equation (9) and (11) (Section 2)

Equation (9): Plugging in $K_t = \left(A_t Z_t^{\frac{\zeta}{\theta}}\right)^{\frac{1}{1-\frac{\zeta}{\theta}}} \bar{K}_t$, $Y_t = \left(A_t Z_t^{\frac{\zeta}{\theta}}\right)^{\frac{1}{1-\frac{\zeta}{\theta}}} \bar{Y}_t$, $C_t = \left(A_t Z_t^{\frac{\zeta}{\theta}}\right)^{\frac{1}{1-\frac{\zeta}{\theta}}} \bar{C}_t$ to (6), we obtain

$$\frac{\left(A_t Z_t^{\frac{\zeta}{\theta}}\right)^{\frac{1}{1-\frac{\zeta}{\theta}}} \bar{K}_t}{\left(A_t Z_t^{\frac{\zeta}{\theta}}\right)^{\frac{1}{1-\frac{\zeta}{\theta}}} \bar{Y}_t} = (A_t Z_t)^{-1} \left(\left(A_t \frac{1}{\left(A_t Z_t^{\frac{\zeta}{\theta}}\right)^{\frac{1}{1-\frac{\zeta}{\theta}}} \bar{Y}_t} \right)^{1-\frac{\theta}{\zeta}} - q_0^{\theta-\zeta} \right).$$

Simplifying terms and pulling out the first term in the last bracket we have

$$\left(A_t Z_t \frac{\bar{K}_t}{\bar{Y}_t}\right)^{\frac{\zeta}{\theta}} = \left(A_t \frac{\bar{L}}{\left(A_t Z_t^{\frac{\zeta}{\theta}}\right)^{\frac{1}{1-\frac{\zeta}{\theta}}} \bar{Y}_t} \right)^{\frac{\zeta}{\theta}-1} \left(1 - q_0^{\theta-\zeta} \left(\frac{A_t \bar{L}_t}{\left(A_t Z_t^{\frac{\zeta}{\theta}}\right)^{\frac{1}{1-\frac{\zeta}{\theta}}} \bar{Y}_t} \right)^{\frac{\theta}{\zeta}-1} \right)^{\frac{\zeta}{\theta}},$$

which after basic manipulations simplifies to

$$\left(\bar{K}_t\right)^{\frac{\zeta}{\theta}} \left(\bar{L}_t\right)^{1-\frac{\zeta}{\theta}} = \bar{Y} \left(1 - q_0^{\theta-\zeta} (A_t Z_t)^{-1} \left(\frac{\bar{L}_t}{\bar{Y}_t}\right)^{\frac{\theta}{\zeta}-1} \right).$$

Equation (11):

We use (6) and normalize variables as above, which gives

$$\bar{K}_t \equiv (A_t Z_t)^{-1} \bar{Y}_t \left(\bar{K}_t, \bar{L}_t\right) \left(\left(A_t \frac{1}{\left(A_t Z_t^{\frac{\zeta}{\theta}}\right)^{\frac{1}{1-\frac{\zeta}{\theta}}} \bar{Y}_t \left(\bar{K}_t, \bar{L}_t\right)} \right)^{1-\frac{\theta}{\zeta}} - q_0^{\theta-\zeta} \right).$$

Let $\phi_t := \left(A_t Z_t^{\frac{\zeta}{\theta}}\right)^{\frac{1}{1-\frac{\zeta}{\theta}}}$. After some simplifications, implicit differentiation of the expression above gives

$$MPK_t = \frac{\zeta}{\theta} \left((A_t Z_t)^{-1} \left(\frac{A_t \bar{L}_t}{\phi_t \bar{Y}_t \left(\bar{K}_t, \bar{L}_t\right)} \right)^{1-\frac{\theta}{\zeta}} - (A_t Z_t)^{-1} q_0^{\theta-\zeta} \frac{\zeta}{\theta} \right)^{-1}.$$

Note that

$$(A_t Z_t)^{-1} \left(\frac{A_t}{\phi_t}\right)^{1-\frac{\theta}{\zeta}} = 1,$$

which obtains (11).

3. Proof of Lemma 1 (Section 3)

Here we prove that the infinite measure in our leading example is indeed a measure. This is a standard result and it can be found, for example, in Billingsley (1995) Theorem 16.9 p. 212 (see also the discussion underlying equation 16.11 on p. 213). For completeness, we include a proof of this fact.

Define a sequence of positive valued functions

$$f_n(q) = \zeta \left(q + \frac{1}{n} \right)^{-\zeta - 1}.$$

Note the following: $f_n \to f$ pointwisely on $\mathbb{R}_{++}$ and it is a monotone sequence $f_n < f_{n+1} < ....$ and f_1 is an L^1 function on $\mathbb{R}_{++}$ ($|f|$ has finite integral). For any Borel subset of $\mathbb{R}_{++}$ define

$$\mu_n(\mathcal{A}) = M \int_{\mathcal{A}} f_n(n)\, dv$$

and define $\mu : B(R_{++}) \to R_+ \cup \{+\infty\}$:

$$\mu(\mathcal{A}) := \lim_{n\to\infty} \mu_n(n).$$

By the monotone convergence theorem applied to $g_n := f_1 - f_n$, since $g_n > g_{n+1} > ...$, the limit exists. The function is defined on the extended real line (the limit may be infinite). Note that whenever μ is finite it coincides with μ as in text, since

$$\lim_{n\to\infty} \mu_n(n) = M \int_{\mathcal{A}} \lim_{n\to\infty} f_n(n)\, dv = M \int_{\mathcal{A}} \zeta q^{-\zeta-1} dv.$$

Trivially, μ_n is a measure by the properties of the Lebesgue integral of integrable functions (f_n is measurable and integrable on $\mathbb{R}_{++}$). However, while f is measurable on Borel subsets of $\mathbb{R}_{++}$, it is not integrable, and the same reasoning does not apply to f. We will show μ is also a measure by invoking the Caratheodory's extension of measure theorem.

Let us introduce some notation first. Let $\mathcal{A} \in 2^{\mathbb{R}_{++}}$ be an arbitrary subset of $\mathbb{R}_{++}$. We will refer to $\pi = \{\mathcal{I}_i : i \in \mathbb{N}\}$ as a covering of $\mathcal{A}$ whenever $\mathcal{A} \subseteq \cup_n^\infty \mathcal{I}_i$, where $\mathcal{I}_i$ is an arbitrarily closed and nonempty interval of the form $[a_i, b_i] \subset \mathbb{R}_{++}$,$0 < a_i < b_i$. We refer to $\mathscr{P}(\mathcal{A})$ as the set of all possible coverings of the set $\mathcal{A}$. Assume it also contains an empty set. Define the volume function associated with a given covering $\pi = \{\mathcal{I}_i : i \in \mathbb{N}\}$ as

$$vol(\pi) = \sum_{i=1}^{\infty} \lim_{n\to\infty} \mu_n(\mathcal{I}_i),$$

Note the following properties of the volume function: 1) it is nonnegative, 2) $vol(\emptyset) = 0$, 3) $vol(\mathbb{R}_{++}) = \infty$, 4) any split of a nonempty and nondegenerate interval to a countable collection of intervals does not change the volume (the property of measure we started with), and 5) it maps from coverings onto $\mathbb{R}_+^{ext}$. Define a candidate outer measure $\mu^* : 2^{\mathbb{R}^2_{++}} \to \mathbb{R}_+ \cup \{+\infty\}$ as

$$\mu^*(\mathcal{A}) = \inf_{\pi \in \mathscr{P}(\mathcal{A})} vol(\pi).$$

We will verify that μ^* satisfies the requirements for an outer measure.

The steps are standard, albeit the argument draws on a particular approach that works best in this context. We must verify the following conditions: i) μ^* is positive-valued and $\mu^*(\emptyset) = 0$, ii) for any collection of sets $\mathcal{A}_i \in 2^{\mathbb{R}_{++}}$, $i = 1, 2, ...$, if $\mathcal{A} \subset \cup_{i=1}^{\infty} \mathcal{A}_i$, then $\mu^*(\mathcal{A}) \leq \sum_{i=1}^{\infty} \mu^*(\mathcal{A}_i)$. The fact that $\mu(\emptyset) = 0$ follows trivially from the empty set being a covering. As for ii, if $\sum_{i=1}^{\infty} \mu^*(\mathcal{A}_i) = +\infty$, we are done. So, assume $\sum_{i=1}^{\infty} \mu^*(\mathcal{A}_i) < +\infty$ and note that $\mathscr{P}(\mathcal{A})$ is non-empty because $\mathcal{A} \subseteq \cup_{i=1}^{\infty} \left(\frac{1}{i}, i\right)$ for any $\mathcal{A} \subseteq \mathbb{R}_{++}$. Let $\mathscr{P}_0(\mathcal{A})$ be a subset of $\mathscr{P}(\mathcal{A})$ defined as follows:

$$\pi \in \mathscr{P}_0(\mathcal{A}) \Leftrightarrow \exists \pi_i \in \mathscr{P}(\mathcal{A}_i), \pi_i = \left\{\mathcal{I}_1^i, \mathcal{I}_2^i, ...\right\}, \text{ s.t. } \pi = \left\{\mathcal{I}_1^1, \mathcal{I}_2^1, ..., \mathcal{I}_1^2, \mathcal{I}_2^2, ..., \mathcal{I}_1^3, \mathcal{I}_2^3 ....\right\}.$$

In words, $\mathscr{P}_0(\mathcal{A})$ is a covering of $\mathcal{A}$ made exclusively of any coverings of sets $\mathcal{A}_i$. It is clear that $\mathscr{P}_0(\mathcal{A})$ is a restricted covering and that it is nonempty because any covering of individual sets must also work for set $\mathcal{A}$, since $\mathcal{A} \subset \cup_{i=1}^{\infty} \mathcal{A}_i$. We now note the following:

$$\mu^*(\mathcal{A}) \leq \inf_{\pi \in \mathscr{P}_0(\mathcal{A})} vol(\pi) = \inf_{\pi \in \mathscr{P}_0(\mathcal{A})} \left\{ \left(\sum_j \left(\sum_i vol\left(\mathcal{I}_i^j\right) \right) \right) \right\} \leq \sum_{i=1}^{\infty} \mu^*(\mathcal{A}_i).$$

The first inequality comes from taking an infimum over a subset of coverings instead of all coverings. The equality is implied by the definition of volume and the way members of $\mathscr{P}_0(\mathcal{A})$ are structured. The last inequality follows from the partition under infimum being selected jointly for all sets instead each set in separation, which can yield a lower value by optimizing on the overlaps. The fact that $\mu^*(\mathcal{A}_i)$ appears on the right-hand side follows from the fact that $\mathscr{P}_0(\mathcal{A})$ is unrestricted as far as the coverings of individual sets $\mathcal{A}_i$ go; that is, $\mathscr{P}(\mathcal{A}_i) \subset \mathscr{P}_0(\mathcal{A})$. The result follows from Theorem 11.2 and Theorem 11.5 in Wheeden and Zygmund (1977). To see that the resulting measure is σ-finite, note that sets $\mathcal{I}_n = [\frac{1}{n}, n]$ are of finite outer measure, with inf covering trivially given by $\mathcal{I}_n$, and note that $\cup_{n=1}^{\infty} \mathcal{I}_n = \mathbb{R}_{++}$. The fact that the measure is infinite follows trivially from the fact that, for example, $\mu_n\left([\frac{1}{n}, 1]\right)$ diverges to infinity with respect to $n \to \infty$. Q.E.D.

4. Proof of Corollary 2 (Section 3)

Consider equation (24) in the proof of Proposition (2). This is the necessary condition for the production function to be Cobb-Douglas that gave rise to (15) that both T_1, T_2 obey. Note that side-by-side integration of this condition gives

$$k(q^*)\,S(q^*) - k(q_0)\,S(q_0) \equiv \frac{1-\alpha}{\alpha}\int_{q_0}^{q^*\left(\frac{w}{r}\right)} k(q)\,dG,$$

which, by equation (15), and the fact that $S(q_0) = 1$, gives

$$k\left(q^*\left(\frac{w}{r}\right)\right) S\left(q^*\left(\frac{w}{r}\right)\right) - C \equiv \frac{1-\alpha}{\alpha}\int_{q_0}^{q^*\left(\frac{w}{r}\right)} k(q)\,dG, \tag{32}$$

Since k is a strictly increasing function, $q^*\left(\frac{w}{r}\right)$ is well-defined and unique. Furthermore, by (1), we know $k\left(q^*\left(\frac{w}{r}\right)\right) = k\left(k^{-1}\left(\frac{w}{r}\right)\right) = \frac{w}{r}$. The above thus boils down to

$$w\frac{L}{Y}\left(\frac{w}{r}\right) - rC \equiv \frac{1-\alpha}{\alpha}\frac{K}{Y}\left(\frac{w}{r}\right),$$

given

$$\frac{L}{Y}\left(\frac{w}{r}\right) \equiv A^{-1}\int_{q^*}^{\infty} g(q)\,dq = A^{-1}S(q^*) \tag{33}$$

and

$$\frac{K}{Y}\left(\frac{w}{r}\right) \equiv A^{-1}\int_{q_0}^{q^*} k(q)\,g(q)\,dq \tag{34}$$

We have now shown that the implied isoquants of the aggregate production function by T_1, T_2 are identical, which proves the first part.

For the last part, divide both sides of the last equation by $\frac{1}{r}\frac{K}{Y}\left(\frac{w}{r}\right)$, and rewrite it as

$$\frac{w\frac{L}{Y}\left(\frac{w}{r}\right)}{r\frac{K}{Y}\left(\frac{w}{r}\right)} \equiv \frac{\frac{1-\alpha}{\alpha}\frac{1}{r}\frac{K}{Y}\left(\frac{w}{r}\right) + rC}{\frac{1}{r}\frac{K}{Y}\left(\frac{w}{r}\right)}.$$

The limit of this expression for $w \to 0$ (any fixed $r > 0$, $C > 0$) gives

$$\frac{w\frac{L}{Y}\left(\frac{w}{r}\right)}{r\frac{K}{Y}\left(\frac{w}{r}\right)} \equiv \frac{\frac{1-\alpha}{\alpha} + \frac{rC}{\frac{1}{r}\frac{K}{Y}\left(\frac{w}{r}\right)}}{1} \to_{w\to\infty} \frac{1-\alpha}{\alpha},$$

since $\frac{1}{r}\frac{K}{Y}\left(\frac{w}{r}\right) \to \infty$ by equation (1) and (34), given $k(q)\,g(q)$ is bounded from below as assumed. Accordingly, relative factor shares converge as stated. Q.E.D.

5. Explicit solution of the capital requirement function (Section 4)

We will calculate the fixed point implied by equation (18) by hand. The equation defines a recursion of the form: $k(q) = \mathcal{R}\left(\kappa_q, \lambda_q; \bar{\kappa}, \bar{\lambda}\right)$, where $\mathcal{R}\left(x, y; \bar{\kappa}, \bar{\lambda}\right) := rx\mathcal{R}(a, b; a, b) + wy$.

For any task q, let us define the average measure of tasks performed by labor as

$$\lambda_q = \int\limits_{q' \in \Lambda_l(q)} 1 d\mu. \tag{35}$$

Observe that Assumption 2 we have the following properties: $\lambda_q = \lambda^0_{i(q)}\mu^\Lambda_q$. Furthermore, note that by independence in Assumption 2, we have

$$\bar{\lambda} = \frac{\int_{q' \in \Lambda_l(q)} \lambda_{q'} d\mu}{\bar{\mu}^\Lambda} = \int\limits_{q' \in \Lambda_l(q)} \lambda^0_{i(q')} d\mu, \tag{36}$$

where

$$\bar{\mu}^\Lambda = \int\limits_{q' \in \Lambda_l(q)} \mu^\Lambda_{q'} d\mu$$

and where these expressions are independent of q.

Consider now the equation for k :

$$k(q) := \int\limits_{q' \in \Lambda_k(q)} rk(q') d\mu + w\lambda^0_{i(q)}\mu^\Lambda_q.$$

We will now perform recursive substitutions along the lines of Figure 5. Due to lengthy expressions, we present the derivation by showing all steps but omitting the terms that do not change from one line to the next. To that end, we use "$*_i$" in place of omitted terms with the exception of the recursion generating integral,

$$\int\limits_{q' \in \Lambda_k(q)} rk(q') d\mu,$$

which we replace by "..." to indicate it is a recursion generator. We will drop the notation "$q' \in \Lambda_k(q)$" and use $\mu(dq')$ and in place $d\mu$ whenever this shortens the expressions.

In the first step, we replace $k(q')$ (which we will do 3 times to show the pattern):

$$k(q) := r \int\limits_{\Lambda_k(q)} \left(\int\limits_{\Lambda_k(q')} rk(q'') \mu(dq'') + w\lambda^0_{i(q')}\mu^\Lambda_{q'} \right) \mu(dq') + *_0,$$

$$k(q) := r^2 \int\limits_{\Lambda_k(q)} \int\limits_{\Lambda_k(q')} k(q'') \mu(dq) \mu(dq') + r \int\limits_{q' \in \Lambda_k(q)} w\lambda^0_{i(q')}\mu^\Lambda_{q'} d\mu + *_0,$$

which yields

$$k\left(q\right) := ... + r\kappa_q w\bar{\lambda}\bar{\mu}^{\Lambda} + *_0,$$

where κ_q is the average measure of tasks when integrated over set $\Lambda_k\left(q\right)$; that is

$$\kappa_q := \int\limits_{q'\in\Lambda_k(q)} \lambda^0_{i(q')}\mu^{\Lambda}_{q'}d\mu$$

Note that this integral must be finite because the original integral (with rk) was finite, and we just substituted it out with an equivalent expression under no arbitrage. We plug in for $k\left(q''\right)$ from the recursion again to obtain another term:

$$k\left(q\right) := r^2 \int\limits_{\Lambda_k(q)} \int\limits_{\Lambda_k(q')} \left(\int\limits_{\Lambda_k(q'')} rk\left(q'''\right)\mu\left(dq'''\right) + w\lambda^0_{i(q''')}\mu^{\Lambda}_{q''} \right) \mu\left(dq''\right)\mu\left(dq'\right) + *_1 + *_0.$$

From this point on we omit the recursion generator altogether as it is now obvious how it generates subsequent terms. Solving the integral above we obtain

$$k\left(q\right) := ... + \int\limits_{q'\in\Lambda_k(q)} \int\limits_{q''\in\Lambda_k(q')} r^2 w\lambda^0_{i(q'')}\mu^{\Lambda}_{q''}d\mu^2 + *_1 + *_0,$$

$$k\left(q\right) := ... + \int\limits_{q'\in\Lambda_k(q)} r^2\kappa_{q'}\mu^{\Lambda}_{q'}w\bar{\lambda}\bar{\mu}^{\Lambda}d\mu + *_1 + *_0,$$

$$k\left(q\right) := ... + r^2\kappa_q\mu^{\Lambda}_q\bar{\kappa}\bar{\mu}^{\Lambda}w\bar{\lambda}\bar{\mu}^{\Lambda} + *_1 + *_0$$

Applying the recursion generator again to obtain another term, a step we omit because it is clear at this point, we obtain another term

$$k\left(q\right) := ... + \int\limits_{q'\in\Lambda_k(q)} \int\limits_{q''\in\Lambda_k(q')} \int\limits_{q'''\in\Lambda_k(q'')} r^3 w\lambda^0_{i(q''')}\mu^{\Lambda}_{q'''}d\mu^3 + *_2 + *_1 + *_0,$$

which after manipulations gives

$$k\left(q\right) := ... + r^3\kappa_q\mu^{\Lambda}_q w\bar{\lambda}\bar{\mu}^{\Lambda}\bar{\kappa}\bar{\mu}^{\Lambda}\bar{\kappa}\bar{\mu}^{\Lambda} + *_2 + *_1 + *_0,$$

$$k\left(q\right) := ... + *_3 + *_2 + *_1 + *_0,$$

Note that we can rewrite the above as

$$k\left(q\right) := *_0 + \left(*_1 + *_2 + *_3 + ...\right)$$

without loss. We next observe that the terms in bracket form a geometric series,

$$k(q) := *_0 + \sum_{i=1}^{\infty} *_i,$$

where

$$\sum_{i=1}^{\infty} *_i = r\kappa_q \mu_q^\Lambda w \bar{\lambda} \bar{\mu}^\Lambda \left(1 + r\bar{\kappa}\bar{\mu}^\Lambda + r^2 \bar{\kappa}^2 \left(\bar{\mu}^\Lambda\right)^2 + ...\right).$$

Note that we can similarly define $\kappa_q = \kappa_{i(q)}^0 \mu_q^\Lambda$ by Assumption 1. It is clear that summing up this series (and assuming it is convergent) gives

$$k(q) := \left(r\kappa_{i(q)}^0 \left(\frac{w\bar{\lambda}}{1 - r\bar{\kappa}\bar{\mu}^\Lambda}\bar{\mu}^\Lambda\right) + \lambda_{i(q)}^0 w\right)\mu_q^\Lambda.$$

The summation is valid as long as $r\bar{\kappa}\bar{\mu}^\Lambda < 1$, which is a condition for a well defined fixed point.

6. Downward projection of complexity (Section 4)

This section shows analytically why downward projection of complexity yields non-generic expressions. The notation is as in Section B.4 above and that section is an integral part of what follows.

We assume equation (19) defines a recursion of the form: $k(q) = \mathcal{R}(\kappa_q, f_q; \kappa_q, f_q)$, where $\mathcal{R}(x, y; a, b) := rx\mathcal{R}(a, b, a, b) + wy$. That is, the original q is inherited all the way down the tree. And, in addition, without the assumption we have $\lambda_q = \lambda_q^0 \mu_q^\Lambda$, $\kappa_q = \kappa_q^0 \mu_q^\Lambda$, and these objects do not average out. This recursion defines a geometric series. Assuming we can apply the summation of the geometric series, which is not a given here as we discuss, we obtain

$$k(q) = \left(r\kappa_q^0 \frac{\lambda_q w}{1 - r\kappa_q^0 \mu_q^\Lambda}\mu_q^\Lambda + \lambda_q^0 w\right)\mu_q^\Lambda.$$

See Section B.4 how this recursion work in the baseline case and in general.

It is clear that the formula with downward project of complexity is more complicated because it involves λ_q and μ_q^Λ and nontrivially links capital requirement to the set theoretic structure of our model. The baseline model nicely separated capital requirement from the set structure of the model.

Consider now $\mu_q^\Lambda = Cq^{\frac{1}{\alpha}}$, as in our baseline model, and let's see how far we can go. Plugging in, we obtain

$$k(q) = \left(r\kappa_q^0 \frac{\lambda_q w}{1 - r\kappa_q^0 C q^{\frac{1}{\alpha}}} C q^{\frac{1}{\alpha}} + \lambda_q^0 w\right) C q^{\frac{1}{\alpha}}.$$

As we can see, if we assume C very small we can partially address the convergence problem of the geometric series (on bounded but arbitrarily large domain). However, note, by doing so, we have now changed the physical environment and make capital defining sets small, which is the problem we tried to avoid by selecting the Pareto distribution as our baseline. But let us proceed. With small C,the relevant function is thus approximately given by $k(q) \approx Cw\lambda_q q^{\frac{1}{\alpha}}$, since term C^2 vanishes.

This expression may seem close to what we need but it still involves λ_q that depends on q in a systematic way, implying that k is connected to the set theoretic structure of the model. This means that the set structure will need to behave in a specific way to yield Cobb-Douglas aggregation. Let us examine it more closely.

We know from the previous results that the function on the right-hand is of constant elasticity with respect to q, and we can calculate that the elasticity of $k(q) \approx Cw\lambda_q q^{\frac{1}{\alpha}}$ is $\frac{1}{\alpha} + \frac{\lambda'(q)}{\lambda(q)} q$,which is a constant if and only if we have $1 - \lambda(q) = 1 + \frac{\hat{C}}{q}$, where $\hat{C}$ is a constant of integration. The fact that we require λ to be increasing in q implies $\hat{C}$ must be negative.[41] Consequently, there must be a lower bound on q to ensure $\lambda(q) > 0$, implying a downward projection of complexity can only apply to the upper part of the domain of complexity. Furthermore, Cobb-Douglas aggregation requires global properties, and so $k(q)$ schedule on the lower part of the domain has to be defined too, and miraculously smoothly connect so that is has the same elasticity α^{-1} everywhere to deliver a Cobb-Douglas function.

Other functional forms do not work either. Consider, for example, $\mu_q^\Lambda = C\lambda_q q^{\frac{1}{\alpha}}$. This seemingly works because λ_q cancels out for the approximate expression obtained above. But this is a one problematic expression because λ_q comes from cost minimization problem of the plant. It is not clear why the physical description of the environment would involve prices that align with the plant minimization problem. Another possibility is a downward projection coming from λ_q only and not the size of the sets, but this makes little difference because it is the presence of λ_q in the expression for $k(q)$ that creates a problems.

Concluding, while we cannot eliminate the possibility of a downward projection of complexity on purely analytic grounds, the analysis shows that it demands a highly non-generic structure.

7. Proof of sufficiency in Proposition 4 (Section 4)

As explained in Part I, Cobb Douglas aggregation obtains if and only if[42]

$$\int_{q_0}^{q^*} r k^e(q)\, \tilde{g}(q)\, dv \equiv \frac{\alpha}{1-\alpha} k^e(q^*)\, \tilde{S}(q^*)\,, \tag{37}$$

which is what we need to show assuming (31) and we can also use (29) (q_0 must be close to zero to have it defined on the entire domain but with this equation Cobb-Douglas will obtain on part of the domain and the limit is unnecessary at this point). To that end, we plug in for $\kappa(w,r) = \tilde{S}\left(k^{e-1}\left(\frac{w}{r}\right)\right)\frac{w}{1-\alpha}$ to (31) and obtain $k^e(q) = \mu_q^\Lambda \tilde{S}\left(k^{e-1}\left(\frac{w}{r}\right)\right)\frac{w}{1-\alpha}$. Next, we use iden-

[41] Decreasing λ leads to nonsensical conclusions in this setup under positive projection of complexity, which we leave for the reader to analyze.

[42] Again, this goes back to equation 23 in the proof of Proposition 2. see comments above why it applies.

tity $\frac{w}{1-\alpha} \equiv \frac{\alpha}{1-\alpha} r k^e \left(k^{e-1}\left(\frac{w}{r}\right)\right) + w$ and replace "$\frac{w}{1-\alpha}$" term in the previous expression, which gives $k^e(q) = \mu_q^\Lambda \tilde{S}(q^*) \left(\frac{\alpha}{1-\alpha} r k^e \left(k^{e-1}\left(\frac{w}{r}\right)\right) + w\right)$. Next, we use equation (29). to replace $k(q)$ on the right hand side of the previous equation, which gives

$$\mu_q^\Lambda \left(\int_{q_0}^{q^*} r k^e(q') \tilde{g} dv + w \tilde{S}\left(q^*\left(\frac{w}{r}\right)\right)\right)$$
$$= \mu_q^\Lambda \left(\tilde{S}\left(q^*\left(\frac{w}{r}\right)\right)\right) \left(\frac{\alpha}{1-\alpha} r k^e \left(k^{e-1}\left(\frac{w}{r}\right)\right) + w\right).$$

After cancellation of strictly positive term ($\mu_q^\Lambda > 0$), and given the fact that $q^* = k^{e-1}\left(\frac{w}{r}\right)$, we are done because have obtained (37).

Online Appendix C: Extensions

This appendix contains extensions of our theory as referenced in text.

1. Tree trunk ring model of complexity distribution (Section 6)

Here we show how our model can be integrated wit R&D models in which ideas are created from sequences of existing ideas, in the spirit of Weitzman (1998) and Jones (2021). This is related to the last example in Section 5.

In the combinatorial growth models tasks are ideas that emerge in the process of selecting the maximum productivity combination our of all or selected combinatorial combinations of existing ideas, and the set of ideas to make new ideas is expanding as new ideas are introduced to the economy. In our context, since we are characterizing tasks underlying a production process that must be completed or else there is no output, the idea of selecting the highest productivity task is moot. We will seek an alternative mechanism that limits the number of possibilities in creation of new ideas and better relate it to models of R&D. That is what most of this section is about. The last part of the model then presents an example of an R&D process that gives rise to a Pareto distribution G.

The basic idea of linking our theory to models along these lines is based on the logic underlying information theory and the asymptotic equipartition property (AEP)—the key result in information theory. AEP shows that a sufficiently long sequence of realizations of an i.i.d. random variable divides the combinatorial space into two sets: a typical set that has an arbitrarily high probability and a non-typical set that has vanishing probability. The typical set comprises the number of combinations that is approximately of order $\left(2^{H(X)}\right)$, where H is the entropy of the i.i.d. random variable that makes the sequence. In this environment we will associate complexity of a new task with $2^{H(X)}$, since on average sequences that make a new idea will be assumed of length n. For an interested reader, the subsequent section provides a detailed discussion of the underlying information theory that here we take for granted.

Complexity as specificity

Let a task be an idea of how to do something useful from the production point of view. Let an idea be comprised of a finite vector $z^n := (z_1, z_2, z_3, ..., z_n) \in \mathbb{N}_+^{\mathrm{n}}$ of other n ideas from some set I referred to as the *alphabet*; that is, $z_i \in I = \{\iota_1, \iota_2 ..., \iota_m\}$, where m is finite. For simplicity, assume that alphabet elements are equiprobable and occur with frequency $p = 1/m$ in any new idea of length n that they are used to create. Since ideas are previously introduced tasks, what we are assuming here is that new ideas are equally likely to be used to make new ideas in the future or be never used again to make new ideas (terminal ideas). The assumption of equiprobable alphabet can be generalized to "*on average* equally likely" but we will not pursue such a generalization here.[43] With this generalization in mind, however, equiprobable alphabet should not be considered as a major shortcoming of the model we are about to develop.

Consider now the set $\mathscr{C}$ of combinatorial allocations of np realizations of each idea from I onto n empty spaces in vector z^n. Note that this is a problem of allocating "balls" into "urns" that can hold only a single ball. The number of such combinations can be calculated as follows:

$$|\mathscr{C}| \approx \begin{pmatrix} n \\ np \end{pmatrix} \times \begin{pmatrix} n-p \\ np \end{pmatrix} \times \begin{pmatrix} n-2p \\ np \end{pmatrix} \times .... \times \begin{pmatrix} n-np \\ np \end{pmatrix}. \tag{38}$$

It turns out that the formula that comes out of the above as a function of n and p characterizes the cardinality of all the vectors z^n that occur with probability 1 as $n \to \infty$—up to a constant of proportionality that will not be relevant in the context of our model. To top it off, all elements from this set are equiprobable, which justifies the assumption that an average new idea contributes on average the same to production of other ideas. The key here is that the cardinality of this set $|\mathscr{C}|$ is not equal to the cardinality of all possible combinatorial combinations (which is 2^n), and it depends on p. These results are a consequence of the law of large numbers, but this requires a lengthy argument that ventures into the theory of large deviations. We discuss it in the next section of this appendix. There, we also prove AEP and show that these combinatorial combinations are the AEP-typical up to a constant.

Applying the Stirling formula ($n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n O(\log n)$) to the above, and noting that the product of the consecutive terms cancels out the numerator of the subsequent expression in the product,

$$\begin{pmatrix} n\,(1-(i-1)\,p) \\ np \end{pmatrix} \begin{pmatrix} n\,(1-ip) \\ np \end{pmatrix} \approx \frac{\sqrt{(1-(i-1)p)}\,((1-(i-1)p))^{S(1-(i-1)p)}}{\left(\sqrt{2\pi np}p^{np}\right)^2},$$

[43]Frequency of large (tail) deviations from the mean decays at an exponential rate as the number of realizations (here S) increases (by Sanov's theorem or Cramer's theorem of theory of large deviations). This can be used to obtain approximately similar expressions to the ones that come out of our model.

we obtain $|\mathscr{C}|^{\frac{1}{n}} \approx \left(\sqrt{2\pi n p} p^{np}\right)^{-1}$. We can further approximate this expression for an arbitrarily large n by omitting the term under the square root, which in logs yields

$$\frac{\log|\mathscr{C}|}{n} \approx -(mp)\log(p) = -\log(p),$$

and hence

$$|\mathscr{C}|^{\frac{1}{n}} \approx p^{-1} = m, \tag{39}$$

since $p = 1/m$. As we discuss in the next section, this approximation is tight in the limit; that is, becomes an equality, but the exponent $1/n$ is necessary to drive the error term implied by the Stirling formula to zero. The true relation is $|\mathscr{C}|^{\frac{1}{n}} = Cm$, for some $C > 1$, because the combinatorial combinations near np realizations must also be take. In our theory the presence of the constant C will be unimportant. The right-hand side, note, does not depend on n, which is convenient.

This derivation shows that $\log|\mathscr{C}|/n$ measures the specificity of an average entry of vector z^n—that is, it is the entropy of its average element. Entropy returns the minimal number of bits that are needed to encode all $\mathscr{C}$ combinations, since $\log_2|\mathscr{C}|$ is a number such $2^{\log_2|\mathscr{C}|}$ is sufficient to label all $|\mathscr{C}|$ elements (natural logarithm is proportional by the change of base formula). For example, if there are 4 combinations possible and $|\mathscr{C}| = 4$, we can fully label these combinations by just two bits because two bits generate a set of 4 unique labels: $\{01, 11, 10, 00\}$. This implies that—on average—information content is $4/n$ bits per item in the vector, or $2^{4/n}$ average combinations. The latter number tells us in a different way how unique or specific vector z^n of normalized length is and this is how we will use it. This basic idea is the foundation of information theory.

We will postulate that the object on the left-hand side in the above equation (specificity) is related our model's notion of complexity via a constant elasticity relationship of the form:

$$q = q_0|\mathscr{C}|^{\frac{1}{n}} = q_0 m, \tag{40}$$

where $q_0 > 0$ is some constant. The definition is motivated by the idea that tasks that were generated from a richer set of possibilities in AEP-typical sense end up being more "complex" for capital to complete, with a linear relationship between information content and complexity. This connection is loose but it makes intuitive sense. As a motivating example, consider a task of coming up with an idea for an economic paper. Such a task is specific in the above sense because the number of sheer possibilities is so large. To employ capital to do this we need to build a machine that will be able to sort out through the vast number of possibilities, which our assumption implies makes it difficult to do. In fact, there is no capital to date that could substitute for labor in completing this task, or in the language of our theory, the capital requirement is nearly infinite relative to the cheap labor provided by a PhD economist.

As we show next, this extension delivers a convenient link between q and the creation of ideas that can be used to endogenize the distribution of tasks in alternative ways to the one discussed in text. We explore one such model based on the Yule process. This is just an example and this bridge can be used to relate our theory to other models that involve combinatorial R&D.

Tree trunk ring growth process

Let the task space be partitioned into a discrete number of subsets of tasks, $F_f = \{T_i\}_{i=1...N_f}$, referred to as creative *families* of tasks. Assume only tasks from the same family can be combined together to create a new task then belonging to the same family and with some constant arrival rate a new task creating a seed for new family. Let there be a small set of *universal tasks* that can be used across all families so that the new family can then grow on its own.

The basic idea here is that of the Yule process. The Yule process is a celebrated result in evolutionary biology and one of the key approaches to rationalize the fact that many processes in nature exhibit power law (tail) behavior. In our view it is an appealing starting point to think about the evolution of ideas. The power law obtains in such an environment because any quantity that grows exponentially and is stopped after exponentially distributed time—which naturally arises here because of the creation of new families—exhibits a power law behavior. Mathematically the key here is that an exponential of an exponentially distributed random variable is Pareto.

To simplify the math, we assume that the process is discrete and new tasks do not join the family immediately but with an epoch delay $\Delta t > 0$, where we will take $\Delta t \to 0$ at the end as an approximation for a continuous process of this kind. The length of the chains of tasks that make a new task—as assumed above—is normalized to some large but fixed value n. (Having it be random would not make any difference as long as it is independent of the number of ideas in the family.) Furthermore, as discussed, with an arrival rate χ newly created tasks branch out and create a new family. The new family takes off thanks to mixing with the universal tasks at first and then grows on its own. Finally, the R&D process that creates new ideas is assumed consistent with the balance growth path in the sense that a constant fraction of economy's resources are devoted to the development of a new task within each family. This leads to an identical exponential growth of tasks within each family.[44] We proceed under this assumption while acknowledging that a nontrivial structure must be behind it. It is beyond the scope of this example to endogenize it and we leave it for future research. This would, for example, require that the creation of ideas are governed by a payoff function that is proportional to the economy's size and the size of the family—assuming constant productivity of a researcher in creating an n chain that makes a new idea.

It is clear that we have obtained the Yule process for the number of members of each family. It is known that this process exhibits power law tail behavior and there is nothing to prove as far as that property goes. However, the distribution of the number of tasks within each family is *not* what we are seeking because it does *not* define the distribution of complexity in the sense of (40). Complexity

[44]It is enough that it is exponential in expectation and there can variation across families. See Reed and Hughes (2002).

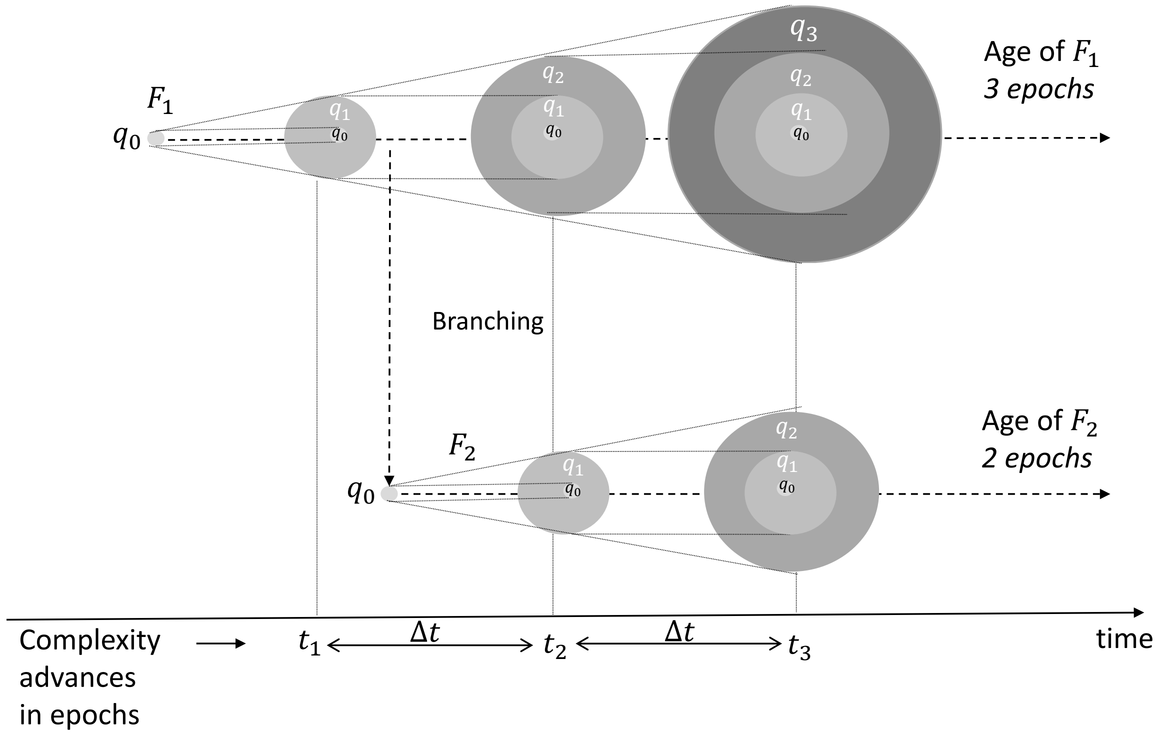


Figure 6: Tree trunk ring model of expansion of complexity.

Notes: The figure shows the growth of two families of tasks and the events of an epoch changes that integrate newly created tasks into the family, raising complexity of task and adding a new ring (shades). Family F_2 is derived from family F_1 via a process of branching (deterministic). We are interested in distribution of these rings at some (late) point in time across all families that exist at this point.

is determined upon the creation of a task, not upon its application.[45] Simple tasks emerge early on when the family is small, and they are simple because their creation involved mixing only a small number of other tasks (ideas). Such process is not a Yule process. We will next show that it also exhibits power law behavior.

Figure 6 illustrates the distribution we want to characterize. The figure shows two families of tasks F_1 and F_2 that evolve over time (see timeline at the bottom). Family F_2 in the figure is derived from family F_1 through the process of branching. The size of each family is represented by the area of the whole circle, which grows exponentially since the inception (for this reason F_2 is smaller). As mentioned, complexity advances in epochs of length Δt as each circle expands exponentially in time, which gives rise to the indicated concentric rings $q_0, q_1, ...$. These rings represent consecutive levels of complexity determined by the advancing number of elapsed epochs. The area of these rings represents the number of tasks of each complexity level. As family F_2 illustrates, the ring growth in newly created families is delayed by the number of epochs the family is younger than its parent family (here F_1). The distribution we are interested in characterizing is the size of the concentric rings $q_0, q_1, ...$ across all families at some distant point in time.

This distribution exhibits power law behavior for the same mathematical reason as for the Yule

[45]Had we assumed specificity as being determined by the current size of the family we would have implied that capital requirement rises for an already invented tasks just because new tasks are invented, which makes little sense

process itself (number of members across families). Intuitively, the number of a given q-level is exponentially distributed because the age of a randomly selected family at any time is exponentially distributed—as it is a result of a Poisson arrival rate with exponential distribution between arrivals. Formally, this result obtains from a simple calculation of the fraction of families that is older than a as a fraction of total number of families. Note that the number of families that are a older is given by the total number of families a periods before relative to the total number at the current moment, which, given the number of families is on average growing at an exponential rate, gives

$$Pr\left(A > a\right) = \frac{\exp\left(\chi\left(t - a\right)\right)}{\exp\left(\chi t\right)} = \exp\left(-\chi a\right).$$

As expected, this is the counter cumulative of the exponential distribution. Since age is proportional to the number of rings, and the size of each consecutive ring (its circumference) is exponential in the age of the family, power law behavior obtains because this implies exponential growth stopped after exponentially distributed time. We summarize this result below and and prove it next.

Proposition 6. *After a sufficiently long period of time, the limit distribution of complexity is* $G\left(q\right) = 1 - \left(\frac{q_0}{q}\right)^{\frac{\chi}{\gamma}}$ *from* $\Delta t \to 0$*, where* $q \geq q_0$.

Proof of Proposition 6:

Consider some period t after a sufficiently long time to allow for application of the law of large numbers. The first observation to make is that at that point the age of each of the very large number of families in epochs is approximately exponentially distributed, with the error being bounded by Δt. Formally, assuming the law of large numbers, we know that the average growth of a large number of families is χ, and so the fraction of families of age a is the number of families a periods of the total number at the current moment, which implies

$$Pr\left(A > a\right) = \frac{\exp\left(\chi\left(t - a\right)\right)}{\exp\left(\chi t\right)} = \exp\left(-\chi a\right), \tag{41}$$

Let $m = F\left(\frac{a}{\Delta t}\right)$,where the F returns the integer floor of any real number. It is clear that $(m-1)\Delta t \leq a < m\Delta t$, which is easy to rearrange to bound on $m\Delta t$ in terms of a. It is clear that for small Δt the error is negligible.

The number of epochs m implied by family's age corresponds to the number of distinct levels of complexity q_i in that family. Formally, let the consecutive levels of complexity up to epoch m after birth be numbered $q_1, q_2, q_3, ...q_m$. Observe that the number of member in each epoch is exponential. Intuitively, each subsequent ring grows on top of the previous one and its circumference is larger. This follows from the fact that the number of members of the family of a age a grows exponentially,

$$|F|_a = \exp\left(\gamma a\right).$$

and growth applies to the initial starting point population $q_0 \exp(\gamma(i-1)\Delta t)$ that builds on top of previous rings; that is

$$q_i = q_0 \exp(\gamma(i-1)\Delta t)\exp(\gamma\Delta t) = q_0 \exp(\gamma i \Delta t).$$

We have now obtained a modified double exponential structure mentioned in text. To see that it exhibits power law behavior, evaluate the probability of drawing a particular level of q_i :

$$Pr(q_i \leq q) = P((q_0 \exp(\gamma i \Delta t)) \leq q).$$

By (41), we observe

$$Pr(q_i \leq q) \approx Pr\left(a \leq \log\left(\left(\frac{q}{q_0}\right)^{\frac{1}{\gamma}}\right)\right) = \exp\left(-\chi \log\left(\frac{q}{q_0}\right)^{\frac{1}{\gamma}}\right),$$

and hence

$$Pr(q_i \leq q) \approx 1 - \left(\frac{q_0}{q}\right)^{\frac{\chi}{\gamma}}.$$

It is clear that the approximation error vanishes as $\Delta t \to 0$ and also averages out across a large number of families by the law of large numbers (we stop the process after a sufficiently long time). We omit the formal argument of bracketing the distribution from above and below by a corresponding epoch point as it is trivial to do so.

Q.E.D.

Entropy as a measure of specificity: background theory

This section provides the background theory underlying the use of formula (38) to motivate our measure of complexity as specificity. We first justify our approach by demonstrating a simpler and a more intuitive setup that counts the sequences in which idea ι occurs exactly np time. We then show this argument generalizes to our model setup and we then discuss its relation to AEP—which is the key result our derivation explicitly uncovers.

Let X_k be a random variable that is 1 if a fixed idea ι occurs in k-th position of the vector $z^n = (z_1, z_2, ..., z_n) \in Z^n$ of a newly minted idea and 0 otherwise. Let $S_n = \sum_{k=1}^{n} x_k$ be the random variable associated with the partial sum of all the appearances of idea ι in the newly minted idea of length n. We will use the notation $x^n = (x_1, x_2, ..., x_n)$, where $x_i \in \{0, 1\}$, to denote a particular realization of the vector as 0 or 1 depending on whether idea ι appears at position $k = 1, 2, ..., n$ in the vector (1 means it appears). Clearly, the probability of $x_i = 1$ is p.

By Chernoff's theorem (see, for example, Billingsley (1995) Theorem 9.3), for any $\varepsilon > 0$, we know that

$$\lim_{n\to\infty}\frac{1}{n}\log Pr\left\{S_n\geq np+n\varepsilon\right\}=-\xi\left(\varepsilon,p\right),$$

where

$$\xi\left(\varepsilon,p\right):=-\left(p+\varepsilon\right)\log\frac{p+\varepsilon}{p}-\left(1-p+\varepsilon\right)\log\frac{1-p-\varepsilon}{p},$$

and hence

$$Pr\left\{S_n\geq n\left(p+\varepsilon\right)\right\}\propto\exp(-n\xi\left(\varepsilon,p\right))\to 0.$$

Applying the same relation to the "no occurrence of idea ι,"

$$S_n^c=\sum_{k=1}^{n}\left(1-X_k^i\right)=n-S_n^i,$$

we also have

$$Pr\left\{n-S_n\geq n\left(\left(1-p\right)+\varepsilon\right)\right\}\propto\exp(-n)\to 0,$$

and hence

$$Pr\left\{S_n\leq n\left(p-\varepsilon\right)\right\}\propto\exp(-n)\to 0.$$

Accordingly, the theory of large deviations tells us that the number of occurrences of a fixed idea ι in the vector z^n within $\pm\varepsilon$ percent of its mean number of occurrences np approaches 1 as $n\to\infty$:

$$Pr\left\{n\left(p-\varepsilon\right)\leq S_n\leq n\left(p+\varepsilon\right)\right\}=1-C\exp(-n)\to_{n\to\infty}1, \tag{42}$$

for some $\varepsilon>0$ and for some constant $C\left(\varepsilon\right)>0$, where $C\left(\varepsilon\right)\to_{\varepsilon\to 0}C>0$. (The last property implies we will not need to worry about this constant depending on ε as we evaluate the limit $\varepsilon\to 0$.)

Let us now define $\mathcal{T}^{\varepsilon}\left(n\right)$ as the set of all z^n such that ι occurs within the range $\pm\varepsilon$ percent of np. (Note that we are switching back to the original vector z^n and random variable Z^n.) We show this is a typical set in the sense that a random idea z^n has probability 1 of being in this set as we take n to the limit—as a corollary from the above result. To show this formally, however, we need to move to a more general probability space of infinite sequences and have a well defined measure under the limit.

To that end, we define a cylinder set of rank n on the space $\Omega:=Z_\infty$ of all possible infinite sequences $z^\infty:=(z_1,z_2,z_3....)$ with respect to some subset $T\subset Z_n$ of the space of finite dimensional sequences Z_n as

$$\Psi_n\left(T\right)=\left\{\omega\in\Omega:\left(z_1\left(\omega\right),z_2\left(\omega\right),...z_n\left(\omega\right)\right)\in T\right\}.$$

(An infinite sequence from Z_∞ such that the first n coordinates are fixed to those from some finite dimensional subset $T \subset Z_n$.) Define the product measure for a cylinder set of any rank as

$$P\left(\Psi_n\left(T\right)\right) = \sum_T p^n.$$

The basic results in probability theory imply that, first, the family of cylinder sets of any rank are a field and the product measure defined above is countably additive probability measure on this field (see, for example, extension of measure in Billingsley (1995) , Section 2). By the Caratheodory's extension of measure theorem, then, the cylinder sets of any rank n can be extended to a $\sigma-$algebra and P has a unique extension to some probability measure $\mathbb{P}$ on that $\sigma-$algebra that coincides with P on the field. As a result, we can say that

$$\forall_{\varepsilon,\delta>0}\exists_{N_{(\varepsilon,\delta)}\in\mathbb{N}}\forall_{n\geq N_{(\varepsilon,\delta)}}\mathbb{P}\left(\Psi_n\left(\mathcal{T}^{\varepsilon}\left(n\right)\right)\right) > 1-\delta,$$

since the probability here pertains to a cylinder set. In probability terms, then, $\mathcal{T}^{\varepsilon}\left(n\right)$ is all that there is for sufficiently large n. Formally, let $\varepsilon_k = \frac{1}{k}$, $\delta_k = \frac{1}{k}$, for $k \in \mathbb{N} :=\{1, 2, 3, ...\}$, and, for any k, pick minimal a sequence $\{n_k\}_k$ such that $n_k \geq N_{(\varepsilon_k,\delta_k)}$ $(n_k = max\{n_{k-1}, N\left(\varepsilon_k, \delta_k\right)\})$, and observe that the above statement implies

$$\mathbb{P}\left(\Psi_{n_k}\left(\mathcal{T}^{\varepsilon_k}\left(n_k\right)\right)\right) \rightarrow_{k\rightarrow\infty} 1.$$

(The number of occurrences of ι in random sequences z^{n_k} within a vanishing radius of $\varepsilon = \frac{1}{k} \rightarrow 0$ percent of the expected number of occurrences $n_k p$ converges to probability 1.)

Since an element is either in the typical set or outside of it, this implies the following decomposition of all sequences z^n to a typical and an a-typical set $\mathcal{A}^{\varepsilon_k}\left(n_k\right) := \Omega \setminus \mathcal{T}^{\varepsilon_k}\left(n_k\right)$:

$$\mathbb{P}\left(\Psi_{n_k}\left(\mathcal{T}^{\varepsilon_k}\left(n_k\right)\right)\right) + \mathbb{P}\left(\Psi_{n_k}\left(\mathcal{A}^{\varepsilon_k}\left(n_k\right)\right)\right) = 1,$$

where we have just established that

$$\mathbb{P}\left(\Psi_{n_k}\left(\mathcal{A}^{\varepsilon_k}\left(n_k\right)\right)\right) \rightarrow_{n\rightarrow\infty} 0.$$

We next characterize the typical set using combinatorics and at the end generalize it to model setup and discuss how it relates to AEP. We also prove AEP.

The number of combinations that have exactly np occurrences of ι among the n entries in the vector z^n is

$$\mathcal{C}\left(n\right) := \left(\begin{array}{c} n \\ np \end{array}\right) = \frac{n!}{(n-np)!np!}.$$

Using the Stirling's approximation ($n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n O(\log n)$), we obtain

$$\mathcal{C}(n) = \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{\sqrt{2\pi n(1-p)}\left(\frac{n(1-p)}{e}\right)^{n(1-p)}\sqrt{2\pi np}\left(\frac{np}{e}\right)^{np}} O(\log n),$$

where O pertains to the "big-O" notation. We next cancel terms, apply $\log_2 \equiv \log$ to both sides, divide by n, and use the law of large numbers to drop terms that are of a lower order than n. This gives

$$\frac{\log \mathcal{C}(n)}{n} \approx \left((1-p)\log(1-p)^{-1} + p\log(p)^{-1}\right) =: H(X)$$

or equivalently (note that $\exp[\frac{O(\log n)}{n}]$ still converges to zero)

$$\mathcal{C}(n)^{\frac{1}{n}} \approx 2^{H(X)}.$$

(Note that the approximation of $\mathcal{C}$ without the exponent would not be valid because of the residual implied by the Stirling formula and this is as far as we can go.)[46]

The above formula measures entropy associated with an average entry "z_i" in vector z^n, which, as the derivation above makes clear, measure the specificity associated with that entry (alternatively, its uniqueness or predictability). To see this, consider the number of bits that it would take to label the entire set of combinations associated with the set $\mathcal{C}$ relative to the number of bits it would take to label all possible combinations, which is 2^n (ones and zeros anywhere in the vector x^n). The number of bits in the case of the set associated with $\mathcal{C}$ is $\log \mathcal{C}$, while the number of bits require to label all combinations is $\log 2^n$ ($\log \equiv \log_2$). Th is is because $2^{\log(.)}$ measures the number of unique labels that a sequence of $\log(.)$ bits can generate. For example, 2 bits, 0 and 1, imply $2^2 = 4$ this generates the set of 4 labels: $01, 10, 11, 00$. Observe now that the entropy $H(X)$ as defined returns the relative number of bits per entry that are needed to label the combinations at hand relative to the number of bits per entry to label all 2^n possible combinations, since

$$\frac{\log \mathcal{C}}{\log 2^n} = \frac{\log \mathcal{C}}{n} \to_{n\to\infty} H(X).$$

(Here note that we can use natural logarithm and the interpretation is the same, since the change of logarithmic base only adds a constant that cancels out. We used a natural logarithm in our model.)

For example, if $p = 1/2$, entropy is maximized and its value is 1, the above calculation implies that $2^n/n$ labels per entry are needed. Intuitively, nothing can be predicted, and so there is no way

[46]It appears that we could exponentiate both sides to obtain the approximation $\mathcal{C} \approx 2^{nH(X)}$ but this approximation is only valid when used under the power $1/n$ for large n (or larger) because of the error implied by Stirling's approximation. One would have to be anyway conscious of this fact to avoid nonsensical conclusions. For example, note that 2^n is the total number combinations possible, but for $p = 1/2$ we have $H(X) = log_2\frac{1}{2} = -1$, implying $\mathcal{C} = 2^n$, and we know from the Binominal coefficient $(n\ np)$ that the total number of combinations as a fraction of the total 2^n goes to zero with n.

to compress information relative to labeling the entire set. However, if $p = 1/10$, entropy is only .47, which implies that a significant compression of information can be achieved.[47] The reason is that in this case mainly zero's will appear, and consecutive zero's can be bunched together (compressed to fixed bit chain) to effectively reduce the information load per average entry. In other words, an average entry is not very specific because is predictable.

The set of combinations identified above is obviously of measure zero, and so it is not yet the typical set $\mathcal{T}^{\varepsilon}(n)$ as defined above. To obtain the typical set, recall, we must add all the combinations within the ε percent deviation of the np occurrences that we used in the derivation (see equation 42). To that end, define

$$\underline{\eta}:=1-\frac{\varepsilon}{p}\leq\eta<\frac{\varepsilon}{p}+1:=\overline{\eta}.$$

It is clear that this range of values for η cover all possibilities from the range that Chernoff's theorem in (42), since $np - n\varepsilon \leq np\eta \leq np + n\varepsilon$. Using the Stirling approximation, the calculation of the number of combinations from this range can obtained by integrating over the continuum of values under an appropriate (conditional) probability measure on this interval—which is well-behaved by the central limit theorem.

We can repeat the above calculations and observe that for any nonzero η we obtain an η-perturbed number of combinations

$$\mathcal{C}^{\eta}(n):=\left(\begin{array}{c} n \\ np\eta \end{array}\right)=\frac{n!}{(n-np\eta)!np\eta!}.$$

which boils down to

$$\mathcal{C}^{\eta}(n)\approx\frac{1}{\sqrt{2\pi n}\sqrt{p\eta(1-p\eta)}\,(1-p\eta)^{n(1-p\eta)}\,(p)^{np\eta}}O(\log n).$$

After integration, then, we obtain

$$\mathcal{C}_{\underline{\eta}}^{\overline{\eta}}(n):=C\int_{\underline{\eta}}^{\overline{\eta}}\mathcal{C}^{\eta}(n)\,\Phi(d\eta)$$

where $C > 1$ is some positive constant and Φ is a conditional probability measure defined on interval $[\underline{\eta},\overline{\eta}]$. (As mentioned, Φ is well-behaved by the central limit theorem, and the constant comes from the fact that we should use a measure with an appropriate scale that corresponds to integer steps in the original formula. That constant has to be at least 1 because we must obtain at least $\mathcal{C}$ combinations from it.).

The function on the right-hand size is continuous in η and by the mean value theorem there exists

[47] For example, if $p = 1/2$, entropy is maximized and its value is 1; that is, $2^n/n$ labels per entry are needed. This is reasonable because nothing can be predicted in this case and no compression of information is possible per entry. However, if $p = 1/10$, the reduction in information load is significant and entropy is only .47.

$\underline{\eta} < \hat{\eta} < \overline{\eta}$ such that

$$\mathcal{C}_{\underline{\eta}}^{\overline{\eta}}(n) = C\frac{\overline{\eta}-\eta}{\overline{\eta}-\underline{\eta}}\mathcal{C}^{\hat{\eta}}(n),$$

where the reason why $\overline{\eta}-\underline{\eta}$ cancels out is because the measure is conditional on this interval. Let us now reuse the previously defined sequence $\varepsilon = 1/k$,n_k, which applied to the above equation (since $\varepsilon = 1/k \to 0$) gives

$$\left(\mathcal{C}_{\underline{\eta}}^{\overline{\eta}}(n_k)\right)^{\frac{1}{n_k}} = \left(\mathcal{C}\mathcal{C}^{\hat{\eta}}(n_k)\right)^{\frac{1}{n_k}} \to_{k\to\infty} 2^{H(X)}.$$

What is the probability of each elements from this set—that is, the combinations in which ι appears exactly $np\eta$ times? Each element is just a permutation of positioning np ι-ideas into n slots in vector z^n, with the probability of its occurrence and so its probability is the same and given by the usual formula:

$$\bar{p}^{\eta}(n) := p^{np\eta}(1-p)^{1-np\eta},$$

Using the previously defined sequence $\varepsilon = 1/k$, n_k, this gives

$$\left(\log \bar{p}^{\eta}(n_k)\right)^{\frac{1}{n_k}} := \left(p^{n_k p\eta}(1-p)^{1-n_k p\eta}\right)^{\frac{1}{n_k}} \to_{k\to\infty} 2^{-H(X)}.$$

As the binominal probability distribution formula also tells us, the probability of the occurrence of such a class of events among all possible events follows the binominal distribution and hence

$$\bar{P}(n) \approx^{*} \hat{C}\int_{\underline{\eta}}^{\overline{\eta}} \mathcal{C}^{\eta}(n)\,\bar{p}^{\eta}(n)\,\Phi(d\eta),$$

where, similarly, we must have $\hat{C} > 1$ because we must obtain a greater number from the integration than the number implied by $\eta = 0$. The approximation sign $\approx^{*}$ means that the formula as stated contains a multiplicative error of order $\log n$ and it is not usable in this form because the error is large (this is the Stirling's formula approximation error). However, it can be used under power $1/n$ as we noted before.

(*)As a verification, we will now show that, in fact, our analysis is consistent and this probability is 1 in the limit. To see this, although the expression on the right-hand side above implies an error of order $O(\log n)$ that makes it too large, we can still apply the mean value theorem to the right-hand side. This means that there exists some $\underline{\eta} < \tilde{\eta} < \overline{\eta}$ (possibly different than $\hat{\eta}$ used before) that approximates the right hand-side integral tightly (for ε small) by

$$\bar{P}(n) = \hat{C}\mathcal{C}^{\tilde{\eta}}(n)\,\bar{p}^{\tilde{\eta}}(n).$$

We will make use of Stirling formula now by raising it to power $1/n$ (alternatively taking log and dividing by n and then exponentiating again as before). This, as before, removes the multiplicative approximation error term "$O(\log n)$" under the limit and the relation

$$\left(\bar{P}\left(n_k\right)\right)^{\frac{1}{n_k}} \approx \left(\hat{C}\mathcal{C}^{\tilde{\eta}}\left(n_k\right)\right)^{\frac{1}{n_k}} \left(\bar{p}^{\tilde{\eta}}\left(n_k\right)\right)^{\frac{1}{n_k}}$$

is without error in the limit $k \to \infty$.

Using the previous expressions and the fact that we can treat $\tilde{\eta}$ the same way as we earlier treated $\hat{\eta}$ under the limit $k \to \infty$ ($\varepsilon = 1/k \to 0$), we obtain the evaluation[48]

$$\lim_{k\to\infty} \left(\bar{P}\left(n_k\right)\right)^{\frac{1}{n_k}} = \lim_{k\to\infty} \left(\hat{C}\mathcal{C}^{\tilde{\eta}}\left(n_k\right)\right)^{\frac{1}{n_k}} \left(\bar{p}^{\tilde{\eta}}\left(n_k\right)\right)^{\frac{1}{n_k}}$$

$$\lim_{k\to\infty} \left(\bar{P}\left(n_k\right)\right)^{\frac{1}{n_k}} = \lim_{k\to\infty} \left(\hat{C}C\right)^{\frac{1}{n_k}} \left(\mathcal{C}\left(n_k\right)\right)^{\frac{1}{n_k}} \left(\bar{p}^{\tilde{\eta}}\left(n_k\right)\right)^{\frac{1}{n_k}}$$

$$\lim_{k\to\infty} \left(\bar{P}\left(n_k\right)\right)^{\frac{1}{n_k}} = \lim_{k\to\infty} \left(\hat{C}C\right)^{\frac{1}{n_k}} 2^{H(X)}2^{-H(X)}$$

$$\lim_{k\to\infty} \left(\bar{P}\left(n_k\right)\right)^{\frac{1}{n_k}} = \lim_{k\to\infty} \left(\hat{C}C\right)^{\frac{1}{n_k}}$$

Since $0 \leq \bar{P}\left(n\right) \leq 1$, the left-hand side either converges to 0 or 1. Since $\hat{C}C > 1$ the right hand can only converge to 1. Accordingly,

$$\bar{P}\left(n_k\right) \to_{k\to\infty} 1.$$

Generalization

The above analysis can be generalized to how we used it in the model; that is, by using (38) instead. It is clear that the Chernoff's theorem can be adopted here as a sufficient condition that each of the ι entries occurs within ε percent radius of its expected number of occurrences np. This is a sufficient condition because the set of combinations that this allows is actually larger then needed and we are seeking the upper bound on the probability. For example, for a finite set of n slots it cannot be the case that, all items ι occur more frequently than the expected value—as we would run of the length of the vector n to fill it. We can thus treat them as independent and by the fact that it enlarges the set replace (42) by

$$\prod_{k=1}^{m} Pr\left\{n\left(p-\varepsilon\right) \leq S_n^{\iota_k} \leq n\left(p+\varepsilon\right)\right\} = \left(1 - C\exp(-n)\right)^m \to_{n\to\infty} 1,$$

where, as before, $S_n^{\iota_m}$ is the random variable associate with the number of times idea ι_m from the alphabet occurs in the sequence z^n. (The exponential term "$\exp(-n)$" converges rapidly, and so this limit is 1 even as m grows in proportion to n. Thecondition can thus accommodate any constant rate of growth of the space of ideas because n can be taken to the limit as fast as we wish.)

[48] Note that in this context the Stirling approximation "$\approx$" is fine enough to imply convergence in the limit.

The extension to perturbed combinations will involve in this case η_ι for each idea (without the last one because it is defined as a dependent variable), and involve multidimensional integration over this set. The integral is cumbersome to write down but conceptually it is no different from what we used above. We can similarly apply the mean value theorem in multiple dimensions—leading to an argument that is qualitatively identical. All this relies on is the continuity of Stirling's approximation, and that does not change. We conclude that a generalization to the setup of our model does not pose any difficulties and follows from the above.

Relation to AEP: A comment

We finish by establishing the connection to AEP, which is the motivating result that we simply uncovered to be able to interpret it. For completeness we prove AEP here and relate it to the analysis above.

Consider $Pr\left(Z_1 = z_1, Z_2 = z_2, ..., Z_n = z_n\right) \equiv Pr\left(z_1, z_2, ..., z_n\right)$: the probability that a particular combination occurs in a sequence of length n (say, as before, ι appearing np times but it can be anything at this point). This probability is

$$Pr\left(z_1, z_2, ..., z_n\right) = \prod_{k=1}^{n} p_k,$$

where p_k is the probability distribution on the set of possibilities $\{\iota_1, \iota_2, ..., \iota_m\}$ (in the example above it is p if ι and $1-p$ otherwise but in general it can be anything).

Taking logs and rearranging, we have

$$-\frac{1}{n}\log Pr\left(z_1, z_2, ..., z_n\right) = -\frac{1}{n}\sum_{k=1}^{n}\log p_k.$$

The right hand size is the statistical mean that converges to the expected value of the random variable $\log Pr\left(Z_i\right)$ in probability (by the weak law of large numbers). Formally, let $\log Pr\left(Z_i\right)$ equal $\log p_k$ whenever realization of random variable $Z_i = z_k$. We thus obtain

$$-\frac{1}{n}\log Pr\left(z_1, z_2, ..., z_n\right) \to^{p} -\mathbb{E}\sum_{i=1}^{n}\log Pr\left(Z_i\right) =: H\left(X\right),$$

or equivalently

$$Pr\left(z_1, z_2, ..., z_n\right) \to^{p} 2^{-nH(X)},$$

which implies

$$\forall_{\delta,\varepsilon>0}\exists_N\forall_{n\geq N} Pr\left\{2^{-n(H(X)+\varepsilon)} \leq Pr\left(X_1, X_2, ..., X_n\right) \leq 2^{-n(H(X)-\varepsilon)}\right\} \geq 1-\delta.$$

We define the AEP-typical set as

$$\hat{\mathcal{T}}^{\varepsilon}(n) := \{x^n : 2^{-n(H(X)+\varepsilon)} \leq Pr(x_1, x_2, ..., x_n) \leq 2^{-n(H(X)-\varepsilon)}\},$$

given the fact that by the previous equation we have (set $\varepsilon = \delta$ in the formula above)

$$\forall_{\varepsilon>0} \exists_N \forall_{n \geq N} Pr\left\{\hat{\mathcal{T}}^{\varepsilon}(n)\right\} \geq 1 - \varepsilon.$$

We have now shown that the previously defined typical set $\mathcal{T}^{\varepsilon}(n)$ using combinatorics and the AEP-typical set $\hat{\mathcal{T}}^{\varepsilon}(n)$ coincide in probability for n sufficiently large—since they both comprise elements that occur with probability 1 in the limit.