9

Artificial Intelligence and Economic Growth

Philippe Aghion, Benjamin F. Jones, and Charles I. Jones

9.1 Introduction

This chapter considers the implications of artificial intelligence for economic growth. Artificial intelligence (AI) can be defined as "the capability of a machine to imitate intelligent human behavior" or "an agent's ability to achieve goals in a wide range of environments."[1] These definitions immediately evoke fundamental economic issues. For example, what happens if AI allows an ever-increasing number of tasks previously performed by human labor to become automated? Artificial intelligence may be deployed in the ordinary production of goods and services, potentially impacting economic growth and income shares. But AI may also change the process by which we create new ideas and technologies, helping to solve complex problems and scaling creative effort. In extreme versions, some observers have argued that AI can become rapidly self-improving, leading to "singularities" that feature unbounded machine intelligence and/or unbounded economic growth in

Philippe Aghion is a professor at the Collège de France and at the London School of Economics. Benjamin F. Jones is the Gordon and Llura Gund Family Professor of Entrepreneurship, professor of strategy, and faculty director of the Kellogg Innovation and Entrepreneurship Initiative at Northwestern University, and a research associate of the National Bureau of Economic Research. Charles I. Jones is the STANCO 25 Professor of Economics at the Graduate School of Business at Stanford University and a research associate of the National Bureau of Economic Research.

We are grateful to Ajay Agrawal, Mohammad Ahmadpoor, Adrien Auclert, Sebastian Di Tella, Patrick Francois, Joshua Gans, Avi Goldfarb, Pete Klenow, Hannes Mahlmberg, Pascual Restrepo, Chris Tonetti, Michael Webb, and participants at the NBER Conference on Artificial Intelligence for helpful discussion and comments. For acknowledgments, sources of research support, and disclosure of the authors' material financial relationships, if any, please see http://www.nber.org/chapters/c14015.ack.

[1] The former definition comes from the Merriam-Webster dictionary, while the latter is from Legg and Hutter (2007).

finite time (Good 1965; Vinge 1993; Kurzweil 2005). Nordhaus (2015) provides a detailed overview and discussion of the prospects for a singularity from the standpoint of economics.

In this chapter, we speculate on how AI may affect the growth process. Our primary goal is to help shape an agenda for future research. To do so, we focus on the following questions:

- If AI increases automation in the production of goods and services, how will it impact economic growth?
- Can we reconcile the advent of AI with the observed constancy in growth rates and capital share over most of the twentieth century? Should we expect such constancy to persist in the twenty-first century?
- Do these answers change when AI and automation are applied to the production of new ideas?
- Can AI drive massive increases in growth rates, or even a singularity, as some observers predict? Under what conditions, and are these conditions plausible?
- How are the links between AI and economic growth modulated by firm-level considerations, including market structure and innovation incentives? How does AI affect the internal organization of firms, and with what implications?

In thinking about these questions, we develop two main themes. First, we model AI as the latest form in a process of automation that has been ongoing for at least 200 years. From the spinning jenny to the steam engine to electricity to computer chips, the automation of aspects of production has been a key feature of economic growth since the Industrial Revolution. This perspective is taken explicitly in two key papers that we build upon: Zeira (1998) and Acemoglu and Restrepo (2016). We view AI as a new form of automation that may allow additional tasks to be automated that previously were thought to be out of reach from automation. These tasks may be nonroutine (to use the language of Autor, Levy, and Murnane [2003]), like self-driving cars, or they may involve high levels of skill, such as legal services, radiology, and some forms of scientific lab-based research. An advantage of this approach is that it allows us to use historical experience on economic growth and automation to discipline our modeling of AI.

A second theme that emerges in our chapter is that the growth consequences of automation and AI may be constrained by Baumol's "cost disease." Baumol (1967) observed that sectors with rapid productivity growth, such as agriculture and even manufacturing today, often see their share of gross domestic product (GDP) decline while those sectors with relatively slow productivity growth—perhaps including many services—experience increases. As a consequence, economic growth may be constrained not by what we do well but rather by what is essential and yet hard to improve. We suggest that combining this feature of growth with automation can yield a rich description of the growth process, including consequences for future growth and income distribution. When applied to a model in which AI automates the production of goods and services, Baumol's insight generates sufficient conditions under which one can get overall balanced growth with a constant capital share that stays well below 100 percent, even with near-complete automation. When applied to a model in which AI automates the production of ideas, these same considerations can prevent explosive growth.[2]

The chapter proceeds as follows. Section 9.2 begins by studying the role of AI in automating the production of goods and services. In section 9.3, we extend AI and automation to the production of new ideas. Section 9.4 then discusses the possibility that AI could lead to superintelligence or even a singularity. In section 9.5, we look at AI and firms, with particular attention to market structure, organization, reallocation, and wage inequality. In section 9.6, we examine sectoral evidence on the evolution of capital shares in tandem with automation. Finally, section 9.7 concludes.

9.2 Artificial Intelligence and Automation of Production

One way of looking at the last 150 years of economic progress is that it is driven by automation. The Industrial Revolution used steam and then electricity to automate many production processes. Relays, transistors, and semiconductors continued this trend. Perhaps artificial intelligence is the next phase of this process rather than a discrete break. It may be a natural progression from autopilots, computer-controlled automobile engines, and MRI machines to self-driving cars and AI radiology reports. While up until recently automation has mainly affected routine or low-skilled tasks, it appears that AI may increasingly automate nonroutine, cognitive tasks performed by high-skill workers.[3] An advantage of this perspective is that it allows us to use historical experience to inform us about the possible future effects of AI.

9.2.1 The Zeira (1998) Model of Automation and Growth

A clear and elegant model of automation is provided by Zeira (1998). In its simplest form, Zeira considers a production function like

$$Y = AX_1^{\alpha_1}X_2^{\alpha_2} \cdot \ldots \cdot X_n^{\alpha_n} \text{ where } \sum_{i=1}^{n}\alpha_i = 1. \tag{1}$$

[2] In the appendix we show that if some steps in the innovation process require human R&D, AI could possibly slow or even end growth by exacerbating business stealing, which in turn discourages human investments in innovation.

[3] Autor, Levy, and Murnane (2003) discuss the effects of traditional software automating routine tasks. Webb et al. (2017) use the text of patent filings to study the different tasks that AI, software, and robotics are best positioned to automate.

While Zeira thought of the X_is as intermediate goods, we follow Acemoglu and Autor (2011) and refer to these as tasks; both interpretations have merit, and we will go back and forth between these interpretations. Tasks that have not yet been automated can be produced one-for-one by labor. Once a task is automated, one unit of capital can be used instead:

$$X_i = \begin{cases} L_i \text{ if not automated} \\ K_i \text{ if automated} \end{cases}. \tag{2}$$

If the aggregate capital K and labor L are assigned to these tasks optimally, the production function can be expressed (up to an unimportant constant) as

$$Y_t = A_t K_t^{\alpha} L_t^{1-\alpha}, \tag{3}$$

where it is now understood that the exponent α reflects the overall share and importance of tasks that have been automated. For the moment, we treat α as a constant and consider comparative statics that increase the share of tasks that get automated.

Next, embed this setup into a standard neoclassical growth model with a constant investment rate; in fact, for the remainder of the chapter this is how we will close the capital/investment side of all our models. The share of factor payments going to capital is given by α and the long-run growth rate of $y \equiv Y/L$ is

$$g_y = \frac{g}{1-\alpha}, \tag{4}$$

where g is the growth rate of A. An increase in automation will therefore increase the capital share α and, because of the multiplier effect associated with capital accumulation, increase the long-run growth rate.

Zeira emphasizes that automation has been going on at least since the Industrial Revolution, and his elegant model helps us to understand that. However, its strong predictions that growth rates and capital shares should be rising with automation go against the famous Kaldor (1961) stylized facts that growth rates and capital shares are relatively stable over time. In particular, this stability is a good characterization of the US economy for the bulk of the twenieth century, for example, see Jones (2016). The Zeira framework, then, needs to be improved so that it is consistent with historical evidence.

Acemoglu and Restrepo (2016) provide one approach to solving this problem. Their rich environment allows for a constant elasticity of substitution (CES) production function and endogenizes the number of tasks as well as automation. In particular, they suppose that research can take two different directions: discovering how to automate an existing task or discovering new tasks that can be used in production. In their setting, a reflects the *fraction* of tasks that have been automated. This leads them to emphasize one possible resolution to the empirical shortcoming of Zeira: perhaps we are inventing new tasks just as quickly as we are automating old tasks. The fraction of tasks that are automated could be constant, leading to a stable capital share and a stable growth rate.

Several other important contributions to this rapidly expanding literature should also be noted. Peretto and Seater (2013) explicitly consider a research technology that allows firms to change the exponent in a Cobb-Douglas production function. While they do not emphasize the link to the Zeira model, with hindsight the connections to that approach to automation are interesting. The model of Hemous and Olsen (2016) is closely related to what follows in the next subsection. They focus on CES production instead of Cobb-Douglas, as we do below, but emphasize the implications of their framework for wage inequality between high-skill and low-skill workers. Agrawal, McHale, and Oettl (2017) incorporate artificial intelligence and the "recombinant growth" of Weitzman (1998) into an innovation-based growth model to show how AI can speed up growth along a transition path.

The next section takes a complementary approach, building on this literature and using the insights of Zeira and automation to understand the structural change associated with Baumol's cost disease.

9.2.2 Automation and Baumol's Cost Disease

The share of agriculture in GDP or employment is falling toward zero. The same is true for manufacturing in many countries of the world. Maybe automation increases the capital share in these sectors and also interacts with nonhomotheticicities in production or consumption to drive the GDP shares toward zero. The aggregate capital share is then a balance of a rising capital share in agriculture/manufacturing/automated goods with a declining GDP share of these goods in the economy.

Looking toward the future, 3D printing techniques and nanotechnology that allow production to start at the molecular or even atomic level could someday automate all manufacturing. Could AI do the same thing in many service sectors? What would economic growth look like in such a world?

This section expands on the Zeira (1998) and Acemoglu and Restrepo (2016) models to develop a framework that is consistent with the large structural changes in the economy. Baumol (1967) observed that rapid productivity growth in some sectors relative to others could result in a "cost disease" in which the slow-growing sectors become increasingly important in the economy. We explore the possibility that automation is the force behind these changes.[4]

[4] The growth literature on this structural transformation emphasizes a range of possible mechanisms, see Kongsamut, Rebelo, and Xie (2001), Ngai and Pissarides (2007), Herrendorf, Rogerson, and Valentinyi (2014), Boppart (2014), and Comin, Lashkari, and Mestieri (2015). The approach we take next has a reduced form that is similar to one of the special cases in Alvarez-Cuadrado, Long, and Poschke (2017).

Model

Gross domestic product is a CES combination of goods with an elasticity of substitution less than one:

$$Y_t = A_t \left(\int_0^1 X_{it}^{\rho} di \right)^{1/\rho} \quad \text{where } \rho > 0, \tag{5}$$

where $A_t = A_0 e^{gt}$ captures standard technological change, which we take to be exogenous for now. Having the elasticity of substitution less than one means that tasks are gross complements. Intuitively, this is a "weak link" production function, where GDP is in some sense limited by the output of the weakest links. Here, these will be the tasks performed by labor, and this structure is the source of the Baumol effect.

As in Zeira, another part of technical change is the automation of production. Goods that have not yet been automated can be produced one-for-one by labor. When a good has been automated, one unit of capital can be used instead:

$$X_{it} = \begin{cases} L_{it} \text{ if not automated} \\ K_{it} \text{ if automated} \end{cases}. \tag{6}$$

This division is stark to keep the model simple. An alternative would be to say that goods are produced with a Cobb-Douglas combination of capital and labor, and when a good is automated, it is produced with a higher exponent on capital.[5]

The remainder of the model is neoclassical:

$$Y_t = C_t + I_t, \tag{7}$$

$$\dot{K}_t = I_t - \delta K_t, \tag{8}$$

$$\int_0^1 K_{it} di = K_t, \tag{9}$$

$$\int_0^1 L_{it} di = L. \tag{10}$$

We assume a fixed endowment of labor for simplicity.

Let β_t be the fraction of goods that that have been automated as of date t. Here, and throughout the chapter, we assume that capital and labor are allocated symmetrically across tasks. Therefore, K_t/β_t units of capital are used in each automated task and $L/(1 - \beta_t)$ units of labor are used on each nonautomated task. The production function can then be written as

$$Y_t = A_t \left[\beta_t \left(\frac{K_t}{\beta_t} \right)^{\rho} + (1 - \beta_t) \left(\frac{L}{1 - \beta_t} \right)^{\rho} \right]^{1/\rho}. \tag{11}$$

[5] A technical condition is required, of course, so that tasks that have been automated are actually produced with capital instead of labor. We assume this condition holds.

Collecting the automation terms simplifies this to

$$Y_t = A_t\left(\beta_t^{1-\rho}K_t^{\rho} + (1-\beta_t)^{1-\rho}L^{\rho}\right)^{1/\rho}. \tag{12}$$

This setup therefore reduces to a particular version of the neoclassical growth model, and the allocation of resources can be decentralized in a standard competitive equilibrium. In this equilibrium, the share of automated goods in GDP equals the share of capital in factor payments:

$$\alpha_{Kt} \equiv \frac{\partial Y_t}{\partial K_t}\frac{K_t}{Y_t} = \beta_t^{1-\rho}A_t^{\rho}\left(\frac{K_t}{Y_t}\right)^{\rho}. \tag{13}$$

Similarly, the share of nonautomated goods in GDP equals the labor share of factor payments:

$$\alpha_{Lt} \equiv \frac{\partial Y_t}{\partial L_t}\frac{L_t}{Y_t} = \beta_t^{1-\rho}A_t^{\rho}\left(\frac{L_t}{Y_t}\right)^{\rho}. \tag{14}$$

Therefore the ratio of automated to nonautomated output—or the ratio of the capital share to the labor share—equals

$$\frac{\alpha_{Kt}}{\alpha_{Lt}} = \left(\frac{\beta_t}{1-\beta_t}\right)^{1-\rho}\left(\frac{K_t}{L_t}\right)^{\rho}. \tag{15}$$

We specified from the beginning that we are interested in the case in which the elasticity of substitution between goods is less than one, so that $\rho < 0$. From equation (15), there are two basic forces that move the capital share (or, equivalently, the share of the economy that is automated). First, an increase in the fraction of goods that are automated, β_t, will increase the share of automated goods in GDP and increase the capital share (holding K/L constant). This is intuitive and repeats the logic of the Zeira model. Second, as K/L rises, the capital share and the value of the automated sector as a share of GDP will decline. Essentially, with an elasticity of substitution less than one, the price effects dominate. The price of automated goods declines relative to the price of nonautomated goods because of capital accumulation. Because demand is relatively inelastic, the expenditure share of these goods declines as well. Automation and Baumol's cost disease are then intimately linked. Perhaps the automation of agriculture and manufacturing leads these sectors to grow rapidly and causes their shares in GDP to decline.[6]

The bottom line is that there is a race between these two forces. As more sectors are automated, β_t increases, and this tends to increase the share of automated goods and capital. But because these automated goods experience faster growth, their price declines, and the low elasticity of substitution means that their shares of GDP also decline.

Following Acemoglu and Restrepo (2016), we could endogenize automation by specifying a technology in which research effort leads goods to

[6] Manuelli and Seshadri (2014) offer a systematic account of the how the tractor gradually replaced the horse in American agriculture between 1910 and 1960.

be automated. But it is relatively clear that depending on exactly how one specifies this technology, $\beta_t/(1 - \beta_t)$ can rise faster or slower than $(K_t/L_t)^\rho$ declines. That is, the result would depend on detailed assumptions related to automation, and currently we do not have adequate knowledge on how to make these assumptions. This is an important direction for future research. For now, however, we treat automation as exogenous and consider what happens when β_t changes in different ways.

Balanced Growth (Asymptotically)

To understand some of these possibilities, notice that the production function in equation (12) is just a special case of a neoclassical production function:

$$(16) \qquad Y_t = A_t F\left(B_t K_t, C_t L_t\right) \text{where } B_t \equiv \beta_t^{(1-\rho)/\rho} \text{ and } C_t \equiv (1-\beta_t)^{(1-\rho)/\rho}.$$

With $\rho < 0$, notice that $\uparrow\beta_t \Rightarrow \downarrow B_t$ and $\uparrow C_t$. That is, automation is equivalent to a combination of labor-augmenting technical change and capital-depleting technical change. This is surprising. One might have thought of automation as somehow capital augmenting. Instead, it is very different: it is labor augmenting and simultaneously *dilutes* the stock of capital. Notice that these conclusions would be reversed if the elasticity of substitution were greater than one; importantly, they rely on $\rho < 0$.

The intuition for this surprising result can be seen by noting that automation has two basic effects. These can be seen most easily by looking back at equation (11). First, capital can be applied to a larger number of tasks, which is a basic capital-augmenting force. However, this also means that a fixed amount of capital is spread more thinly, a capital-depleting effect. When the tasks are substitutes ($\rho > 0$), the augmenting effect dominates and automation is capital augmenting. However, when tasks are complements ($\rho < 0$), the depletion effect dominates and automation is capital depleting. Notice that for labor, the opposite forces are at work: automation concentrates a given quantity of labor onto a smaller number of tasks and hence is labor augmenting when $\rho < 0$.[7]

This opens up one possibility that we will explore next: what happens if the evolution of β_t is such that C_t grows at a constant exponential rate? This can occur if $1 - \beta_t$ falls at a constant exponential rate toward zero, meaning that $\beta_t \to 1$ in the limit and the economy gets ever closer to full automation (but never quite reaches that point). The logic of the neoclassical growth model suggests that this could produce a balanced growth path with constant factor shares, at least in the limit. (This requires A_t to be constant.)

In particular, we want to consider an exogenous time path for the fraction

[7] In order for automation to increase output, we require a technical condition: $(K/\beta)^\rho < [L/(1 - \beta)]^\rho$. For $\rho < 0$, this requires $K/\beta > L/1 - \beta$. That is, the amount of capital that we allocate to each task must exceed the amount of labor we allocate to each task. Automation raises output by allowing us to use our plentiful capital on more of the tasks performed by relatively scarce labor.

of tasks that are automated, β_t, such that $\beta_t \rightarrow 1$ but in a way that C_t grows at a constant exponential rate. This turns out to be straightfoward. Let $\gamma_t \equiv 1 - \beta_t$, so that $C_t = \gamma_t^{(1-\rho)/\rho}$. Because the exponent is negative ($\rho < 0$), if γ falls at a constant exponential rate, C_t will grow at a constant exponential rate. This occurs if $\dot{\beta}_t$, $= \theta(1 - \beta_t)$, implying that $g_\gamma = -\theta$. Intuitively, a constant fraction, θ, of the tasks that have not yet been automated become automated each period.

Figure 9.1 shows that this example can produce steady exponential growth. We begin in year 0 with none of the goods being automated, and then have a constant fraction of the remainder being automated each year. There is obviously enormous structural change underlying—and generating—the stable exponential growth of GDP in this case. The capital share of factor payments begins at zero and then rises gradually over time, eventually asymptoting to a value around one-third. Even though an ever-vanishing fraction of the economy has not yet been automated, so labor has less and less to do. The fact that automated goods are produced with cheap capital combined with an elasticity of substitution less than one means that the automated share of GDP remains at one-third and labor still earns around two-thirds of GDP asymptotically. This is a consequence of the Baumol force: the labor tasks are the "weak links" that are essential and yet expensive, and this keeps the labor share elevated.[8]

Along such a path, however, sectors like agriculture and manufacturing exhibit a structural transformation. For example, let sectors on the interval [0,1/3] denote agriculture and the automated portion of manufacturing as of some year, such as 1990. These sectors experience a declining share of GDP over time, as their prices fall rapidly. The automated share of the economy will be constant only because new goods are becoming automated.

The analysis so far requires A_t to be constant, so that the only form of technical change is automation. This seems too extreme: surely technical progress is not only about substituting machines for labor, but also about creating better machines. This can be incorporated in the following way. Suppose A_t is *capital-augmenting* rather than Hicks-neutral, so that the production function in equation (16) becomes $Y_t = F(A_t B_t K_t, C_t L_t)$. In this case, one could get a balanced growth path (BGP) if A_t rises at precisely the rate that B_t declines, so that technological change is essentially purely labor-augmenting on net: better computers would decrease the capital share at precisely the rate that automation raises it, leading to balanced growth. At first, this seems like a knife-edge result that would be unlikely in practice. However, the logic of this example is somewhat related to the model in Grossman et al. (2017); that paper presents an environment in which it is optimal to have something similar to this occur. So perhaps this alternative

[8] The neoclassical outcome here requires that θ not be too large (e.g., relative to the exogenous investment rate). If θ is sufficiently high, the capital share can asymptote to one and the model becomes "AK." We are grateful to Pascual Restrepo for working this out.

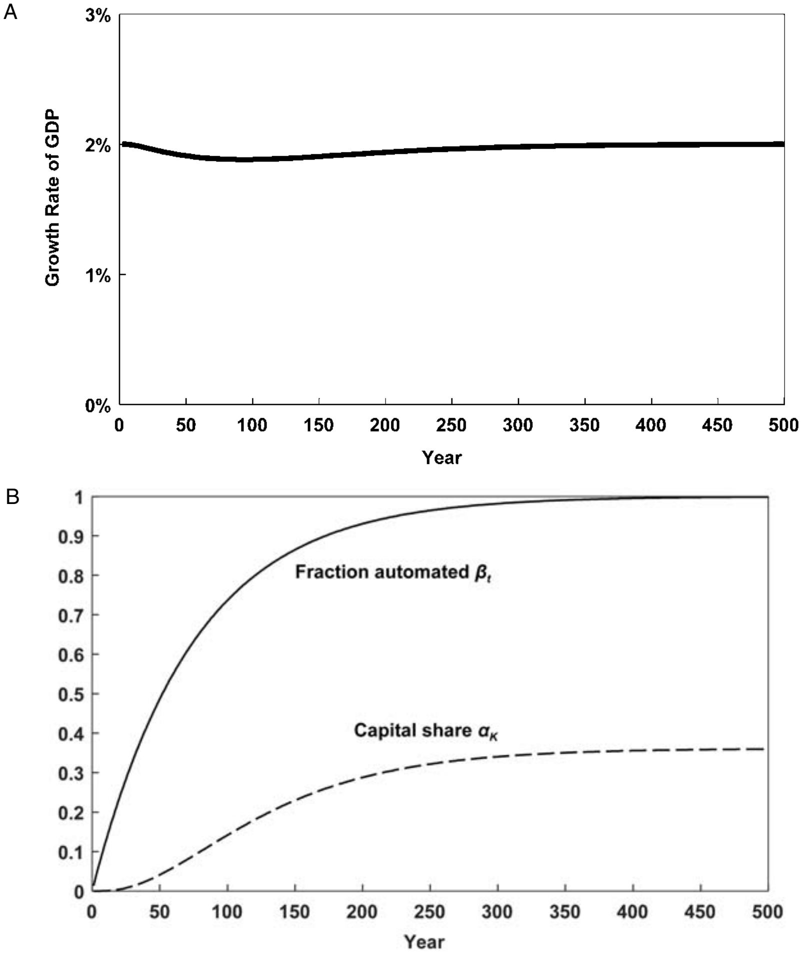

Fig. 9.1 Automation and asymptotic balanced growth. *A*, the growth rate of GDP over time; *B*, automation and the capital share

Note: This simulation assumes $\rho < 0$ and that a constant fraction of the tasks that have not yet been automated become automated each year. Therefore $C_t \equiv (1 - \beta)^{(1-\rho)/\rho}$ grows at a constant exponential rate (2 percent per year in this example), leading to an asymptotic balanced growth path (BGP). The share of tasks that are automated approaches 100 percent in the limit. Interestingly, the capital share of factor payments (and the share of automated goods in GDP) remains bounded, in this case at a value around one-third. With a constant investment rate of $\bar{s}$, the limiting value of the capital share is $(\bar{s}/g_Y + \delta)^\rho$.

approach could be given good microfoundations. We leave this possibility to future research.

Constant Factor Shares

Another interesting case worth considering is under what conditions can this model produce factor shares that are constant over time? Taking logs and derivatives of equation (15), the capital share will be constant if and only if

$$(17) \qquad g_{\beta t} = \left(1-\beta_t\right)\left(\frac{-\rho}{1-\rho}\right)g_{kt},$$

where g_{kt} is the growth rate of $k \equiv K/L$. This is very much a knife-edge condition. It requires the growth rate of β_t to slow over time at just the right rate as more and more goods get automated.

Figure 9.2 shows an example with this feature, in an otherwise neoclassical model with exogenous growth in A_t at 2 percent per year. That is, unlike the previous section, we allow other forms of technological change to make tractors and computers better over time, in addition to allowing automation. In this simulation, automation proceeds at just the right rate so as to keep the capital share constant for the first 150 years. After that time, we simply assume that β_t is constant and automation stops, so as to show what happens in that case as well.

The perhaps surprising result in this example is that the constant factor shares occur while the growth rate of GDP rises at an increasing rate. From the earlier simulation in figure 9.1, one might have inferred that a constant capital share would be associated with declining growth. However, this is not the case and instead growth rates increase. The key to the explanation is to note that with some algebra, we can show that the constant factor share case requires

$$(18) \qquad g_{Yt} = g_A + \beta_t g_{Kt}.$$

First, consider the case with $g_A = 0$. We know that a true balanced growth path requires $g_Y = g_K$. This can occur in only two ways if $g_A = 0$: either $\beta_t = 1$ or $g_Y = g_K = 0$ if $\beta_t < 1$. The first case is the one that we explored in the previous example back in figure 9.1. The second case shows that if $g_A = 0$, then constant factor shares will be associated with zero exponential growth.

Now we can see the reconciliation between figures 9.1 and 9.2. In the absence of $g_A > 0$, the growth rate of the economy would fall to zero. Introducing $g_A > 0$ with constant factor shares *does* increases the growth rate. To see why growth has to accelerate, equation (18) is again useful. If growth were balanced, then $g_Y = g_K$. But then the rise in β_t would tend to raise g_Y and g_K. This is why growth accelerates.

Regime Switching

A final simulation shown in figure 9.3 combines aspects of the two previous simulations to produce results closer in spirit to our observed data, albeit in a highly stylized way. We assume that automation alternates between two regimes. The first is like figure 9.1, in which a constant fraction of the remaining tasks are automated each year, tending to raise the capital share and produce high growth. In the second, β_t is constant and no new automation occurs. In both regimes, A_t grows at a constant rate of 0.4 percent per

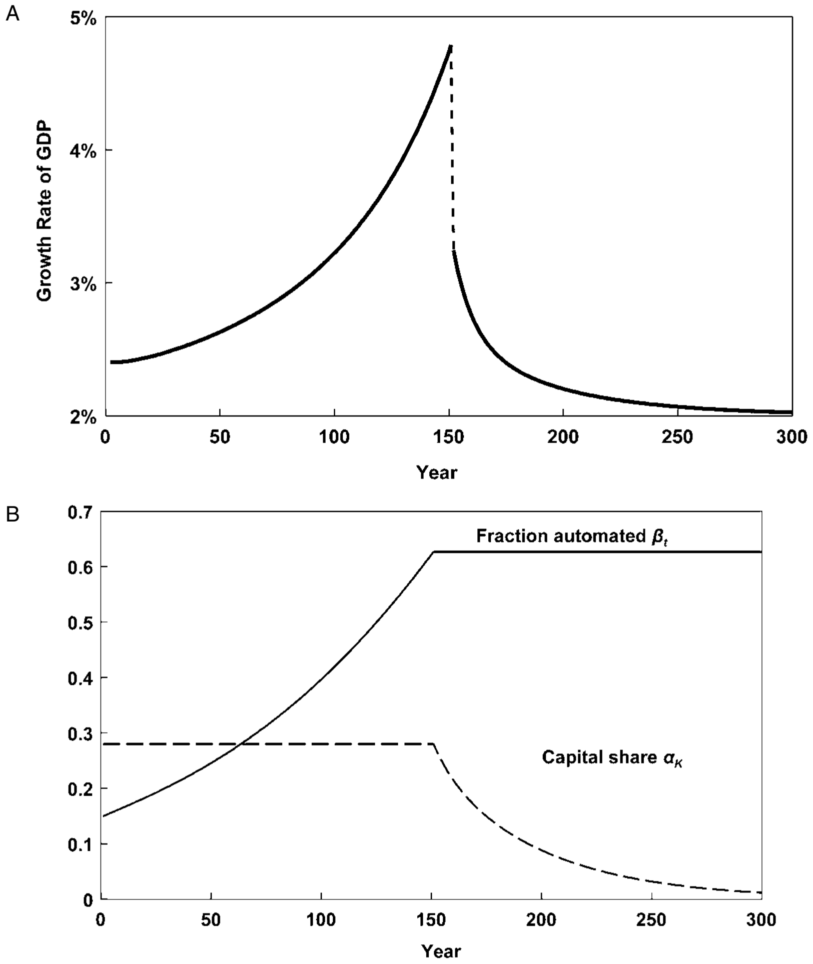

Fig. 9.2 Automation with a constant capital share. *A*, the growth rate of GDP over time; *B*, automation and the capital share

Note: This simulation assumes $\rho < 0$ and sets β_t so that the capital share is constant between year 0 and year 150. After year 150, we assume β_t stays at its constant value; A_t is assumed to grow at a constant rate of 2 percent per year throughout.

year, so that even when the fraction of tasks being automated is stagnant, the nature of automation is improving, which tends to depress the capital share. Regimes last for thirty years. Period 100 is highlighted with a black circle. At this point in time, the capital share is relatively high and growth is relatively low.

By playing with parameter values, including the growth rate of A_t and β_t, it is possible to get a wide range of outcomes. For example, the fact that

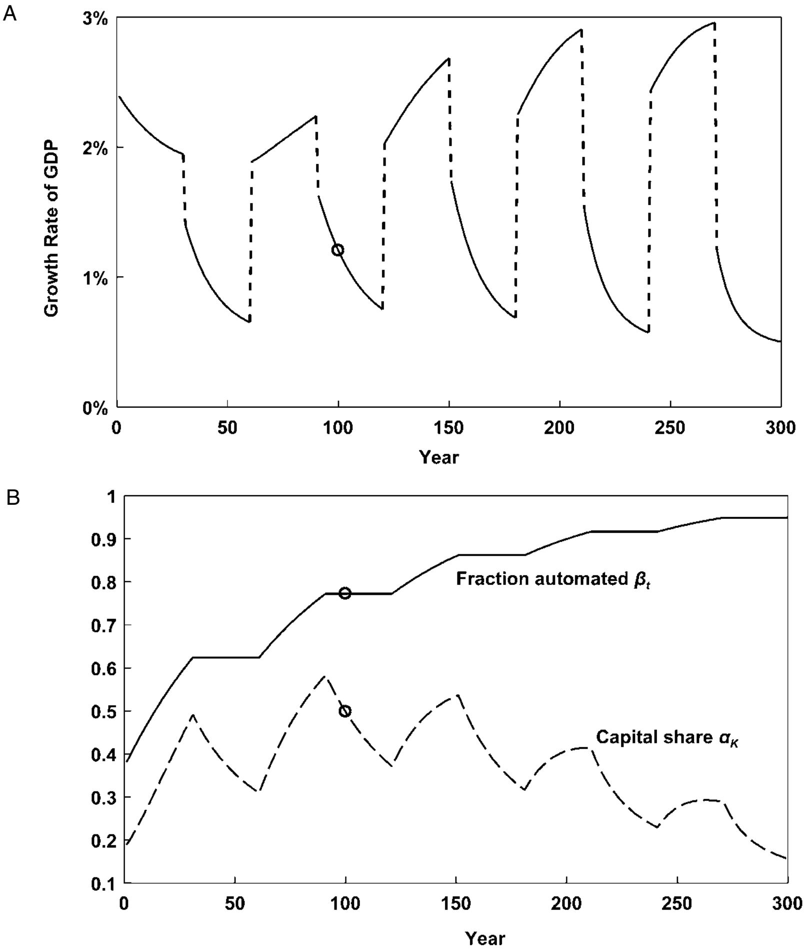

Fig. 9.3 Intermittent automation to match data? *A*, the growth rate of GDP over time; *B*, automation and the capital share

Note: This simulation combines aspects of the two previous simulations to produce results closer in spirit to our observed data. We assume that automation alternates between two regimes. In the first, a constant fraction of the remaining tasks are automated each year. In the second, β_t is constant and no new automation occurs. In both regimes, A_t grows at a constant rate of 0.4 percent per year. Regimes last for thirty years. Period 100 is highlighted with a black circle. At this point in time, the capital share is relatively high and growth is relatively low.

the capital share in the future is lower than in period 100 instead of higher can be reversed.

Summing Up

Automation—an increase in β_t—can be viewed as a "twist" of the capital- and labor-augmenting terms in a neoclassical production function. From

Uzawa's famous theorem, since we do not in general have purely labor-augmenting technical change, this setting will not lead to balanced growth. In this particular application (e.g., with $\rho < 0$), either the capital share or the growth rate of GDP will tend to increase over time, and sometimes both. We showed one special case in which all tasks are ultimately automated that produced balanced growth in the limit with a constant capital share less than 100 percent. A shortcoming of this case is that it requires automation to be the *only* form of technological change. If, instead, the nature of automation itself improves over time—consider the plow, then the tractor, then the combine-harvester, then GPS tracking—then the model is best thought of as featuring both automation and something like improvements in A_t. In this case, one would generally expect growth not to be balanced. However, a combination of periods of automation followed by periods of respite, like that shown in figure 9.3 does seem capable of producing dynamics at least superficially similar to what we have seen in the United States in recent years: a period of a high capital share with relatively slow economic growth.

9.3 Artificial Intelligence in the Idea Production Function

In the previous section, we examined the implications of introducing AI in the production function for goods and services. But what if the tasks of the innovation process themselves can be automated? How would AI interact with the production of new ideas? In this section, we introduce AI in the production technology for new ideas and look at how AI can affect growth through this channel.

A moment of introspection into our own research process reveals many ways in which automation can matter for the production of ideas. Research tasks that have benefited from automation and technological change include typing and distributing our papers, obtaining research materials and data (e.g., from libraries), ordering supplies, analyzing data, solving math problems, and computing equilibrium outcomes. Beyond economics, other examples include carrying out experiments, sequencing genomes, exploring various chemical reactions and materials. In other words, applying the same task-based model to the idea production function and considering the automation of research tasks seems relevant.

To keep things simple, suppose the production function for goods and services just uses labor and ideas:

$$Y_t = A_t L_t. \tag{19}$$

But suppose that various tasks are used to make new ideas according to

$$\dot{A}_t = A_t^{\phi} \left(\int_0^1 X_{it}^{\rho} di \right)^{1/\rho} \text{ where } \rho < 0. \tag{20}$$

Assuming some fraction β_t of tasks have been automated—using a similar setup to that in section 9.2—the idea production function can be expressed as

$$\dot{A}_t = A_t^{\phi}\Big((B_tK_t)^{\rho} + (C_tS_t)^{\rho}\Big)^{1/\rho} \equiv A_t^{\phi}F\big(B_tK_t, C_tS_t\big), \tag{21}$$

where S_t is the research labor used to make ideas, and B_t and C_t are defined as before, namely, $B_t \equiv \beta_t^{(1-\rho)\rho}$ and $C_t \equiv (1-\beta_t)^{(1-\rho)/\rho}$.

Several observations then follow from this setup. First, consider the case in which β_t is constant at some value but then increases to a higher value (recall that this leads to a one-time decrease in B_t and increase in C_t). The idea production function can then be written as

$$\begin{aligned} \dot{A}_t &= A_t^{\phi}S_tF\left(\frac{BK_t}{S_t}, C\right) \\ &\sim A_t^{\phi}CS_t, \end{aligned} \tag{22}$$

where the "~" notation means "is asymptotically proportional to." The second line follows if K_t/S_t is growing over time (i.e., if there is economic growth) and if the elasticity of substitution in $F(\cdot)$ is less than one, which we have assumed. In that case, the CES function is bounded by its scarcest argument, in this case researchers. Automation then essentially produces a level effect but leaves the long-run growth rate of the economy unchanged if $\phi < 1$. Alternatively, if $\phi = 1$—the classic endogenous growth case—then automation raises long-run growth.

Next, consider this same case of a one-time increase in β, but suppose the elasticity of substitution in $F(\cdot)$ equals one, so that $F(\cdot)$ is Cobb-Douglas. In this case, as in the Zeira model, it is easy to show that a one-time increase in automation will raise the long-run growth rate. Essentially, an accumulable factor in production (capital) becomes permanently more important, and this leads to a multiplier effect that raises growth.

Third, suppose now that the elasticity of substitution is greater than one. In this case, the argument given before reverses, and now the CES function asymptotically looks like the plentiful factor, in this case K_t. The model will then deliver explosive growth under fairly general conditions, with incomes becoming infinite in finite time.[9] But this is true even *without* any automation. Essentially, in this case researchers are not a necessary input and so standard capital accumulation is enough to generate explosive growth. This is one reason why the case of $\rho < 1$—that is, an elasticity of substitution less than one—is the natural case to consider. We focus on this case for the remainder of this section.

[9] A closely related case is examined explicitly in the discussion surrounding equation (27) below.

9.3.1 Continuous Automation

We can now consider the special case in which automation is such that the newly automated tasks constitute a constant fraction, q, of the tasks that have not yet been automated. Recall that this was the case that delivered a balanced growth path back in the Balanced Growth section

In This Case, $B_t \rightarrow 1$ and $(\dot{C}_t/C_t) \rightarrow g_c = -[(1-\rho)/\rho]\cdot\theta > 0$ Asymptotically

The same logic that gave us equation (22) now implies that

$$\begin{aligned}\dot{A}_t &= A_t^{\phi} C_t S_t F\left(\frac{B_t K_t}{C_t S_t}, 1\right) \\ &\sim A_t^{\phi} C_t S_t,\end{aligned} \tag{23}$$

where the second line holds as long as $BK/CS \rightarrow \infty$, which holds for a large class of parameter values.[10]

This reduces to the Jones (1995) kind of setup, except that now "effective" research grows faster than the population because of AI. Dividing both sides of the last expression by A_t gives

$$\frac{\dot{A}_t}{A_t} = \frac{C_t S_t}{A_t^{1-\phi}}. \tag{24}$$

In order for the left-hand side to be constant, we require that the numerator and denominator on the right side grow at the same rate, which then implies

$$g_A = \frac{g_C + g_S}{1-\phi}. \tag{25}$$

In Jones (1995), the expression was the same except $g_C = 0$. In that case, the growth rate of the economy is proportional to the growth rate of researchers (and ultimately, the population). Here, automation adds a second term and raises the growth rate: we can have exponential growth in research effort in the idea production function not only because of growth in the actual number of people, but also as a result of the automation of research implied by AI. Put another way, even with a constant number of researchers, the number of researchers per task $S/(1-\beta_t)$ can grow exponentially: the fixed number of researchers is increasingly concentrated onto an exponentially declining number of tasks.[11]

[10] Since $B \rightarrow 1$, we just require that $g_k > g_c$. This will hold—see below—for example if $\phi > 0$.

[11] Substituting in for other solutions, the long-run growth rate of the economy is $g_y = \{-[(1-\rho)/\rho]\cdot\theta + n\}/(1-\phi)$, where n is the rate of population growth.

9.4 Singularities

To this point, we have considered the effects of gradual automation in the goods and idea production functions and shown how that can potentially raise the growth rate of the economy. However, many observers have suggested that AI opens the door to something more extreme—a "technological singularity" where growth rates will explode. John Von Neumann is often cited as first suggesting a coming singularity in technology (Danaylov 2012). I. J. Good and Vernor Vinge have suggested the possibility of a self-improving AI that will quickly outpace human thought, leading to an "intelligence explosion" associated with infinite intelligence in finite time (Good 1965; Vinge 1993). Ray Kurzweil in *The Singularity is Near* also argues for a coming intelligence explosion through nonbiological intelligence (Kurzweil 2005) and, based on these ideas, cofounded Singularity University with funding from prominent organizations like Google and Genentech.

In this section, we consider singularity scenarios in light of the production functions for both goods and ideas. Whereas standard growth theory is concerned with matching the Kaldor facts, including constant growth rates, here we consider circumstances in which growth rates may increase rapidly over time. To do so, and to speak in an organized way to the various ideas that borrow the phrase "technological singularity," we can characterize two types of growth regimes that depart from steady-state growth. In particular, we can imagine:

- a "Type I" growth explosion, where growth rates increase without bound but remain finite at any point in time; and
- a "Type II" growth explosion, where infinite output is achieved in finite time.

Both concepts appear in the singularity community. While it is common for writers to predict the singularity date (often just a few decades away), writers differ on whether the proposed date records the transition to the new growth regime of Type I or an actual singularity occurring of Type II.[12]

To proceed, we now consider examples of how the advent of AI could drive growth explosions. The basic finding is that complete automation of tasks by an AI can naturally lead to the growth explosion scenarios above. However, interestingly, one can even produce a singularity without relying on complete automation, and one can do it without relying on an intelligence explosion per se. Further below, we will consider several possible objections to these examples.

[12] Vinge (1993), for example, appears to be predicting a Type II explosion, a case that has been examined mathematically by Solomonoff (1985), Yudkowsky (2013), and others. Kurzweil (2005), by contrast, who argues that the singularity will come around the year 2045, appears to be expecting a Type I event.

9.4.1 Examples of Technological Singularities

We provide four examples. The first two examples take our previous models to the extreme and consider what happens if everything can be automated—that is, if people can be replaced by AI in all tasks. The third example demonstrates a singularity through increased automation but without relying on complete automation. The final example looks directly at "superintelligence" as a route to a singularity.

Example 1: Automation of Goods Production

The Type I case can emerge with full automation in the production for goods. This is the well-known case of an AK model with ongoing technological progress. In particular, take the model of section 9.2, but assume that *all* tasks are automated as of some date t_0. The production function is thereafter $Y_t = A_t K_t$ and growth rates themselves grow exponentially with A_t. Ongoing productivity growth—for example, through the discovery of new ideas—would then produce ever-accelerating growth rates over time. Specifically, with a standard capital accumulation specification ($\dot{K}_t = \bar{s}Y_t - \delta K_t$) and technological progress proceeding at rate g, the growth rate of output becomes

$$g_Y = g + \bar{s}A_t - \delta, \tag{26}$$

which grows exponentially with A_t.

Example 2: Automation of Ideas Production

An even stronger version of this acceleration occurs if the automation applies to the idea production function instead of (or in addition to) the goods production function. In fact, one can show that there is a mathematical singularity: a Type II event where incomes essentially become infinite in a finite amount of time.

To see this, consider the model of section 9.3. Once all tasks can be automated, that is, once AI replaces all people in the idea production function, the production of new ideas is given by

$$\dot{A}_t = K_t A_t^{\phi}. \tag{27}$$

With $\phi > 0$, this differential equation is "more than linear." As we discuss next, growth rates will explode so fast that incomes become infinite in finite time.

The basic intuition for this result comes from noting that this model is essentially a two-dimensional version of the differential equation $\dot{A}_t = A_t^{1+\phi}$ (e.g., replacing the K with an A in equation [27]). This differential equation can be solved using standard methods to give

$$A_t = \left(\frac{1}{A_0^{-\phi} - \phi t} \right)^{1/\phi}. \tag{28}$$

And it is easy to see from this solution that $A(t)$ exceeds any finite value before date $t^* = (1/\phi A_0^{\phi})$. This is a singularity.

For the two dimensional system with capital in equation (27), the argument is slightly more complicated but follows this same logic. The system of differential equations is equation (27) together with the capital accumulation equation ($\dot{K}_t = \bar{s}Y_t - \mathrm{d}K_t$, where $Y_t = A_t L$). Writing these in growth rates gives

$$\frac{\dot{A}_t}{A_t} = \frac{K_t}{A_t} \cdot A_t^{\phi}, \tag{29}$$

$$\frac{\dot{K}_t}{K_t} = \bar{s}L\frac{A_t}{K_t} - \delta. \tag{30}$$

First, we show that $(\dot{A}_t/A_t) > (\dot{K}_t/K_t)$. To see why, suppose they were equal. Then equation (30) implies that $(\dot{K}_t/K_t)$ is constant, but equation (29) would then imply that $(\dot{A}_t/A_t)$ is accelerating, which contradicts our original assumption that the growth rates were equal. So it must be that $(\dot{A}_t/A_t) > (\dot{K}_t/K_t)$.[13] Notice that from the capital accumulation equation, this means that the growth rate of capital is rising over time, and then the idea growth rate equation means that the growth rate of ideas is rising over time as well. Both growth rates are rising. The only question is whether they rise sufficiently fast to deliver a singularity.

To see why the answer is yes, set $\delta = 0$ and $\bar{s}L = 1$ to simplify the algebra. Now multiply the two growth rate equations together to get

$$\frac{\dot{A}_t}{A_t} \cdot \frac{\dot{K}_t}{K_t} = A_t^{\phi}. \tag{31}$$

We have shown that $(\dot{A}_t/A_t) > (\dot{K}_t/K_t)$, so combining this with equation (31) yields

$$\left(\frac{\dot{A}_t}{A_t}\right)^2 > A_t^{\phi}, \tag{32}$$

implying that

$$\frac{\dot{A}_t}{A_t} > A_t^{\phi/2}. \tag{33}$$

That is, the growth rate of A grows at least as fast as $A_t^{\phi/2}$. But we know from the analysis of the simple differential equation given earlier—see equation (28)—that even if equation (33) held with equality, this would be enough to deliver the singularity. Because A grows faster than that, it also exhibits a singularity.

Because ideas are nonrival, the overall economy is characterized by increasing returns, à la Romer (1990). Once the production of ideas is fully

[13] It is easy to rule out the opposite case of $(\dot{A}_t/A_t) < (\dot{K}_t/K_t)$.

automated, this increasing returns applies to "accumulable factors," which then leads to a Type II growth explosion, that is, a mathematical singularity.

Example 3: Singularities without Complete Automation

The above examples consider complete automation of goods production (Example 1) and ideas production (Example 2). With the CES case and an elasticity of substitution less than one, we require that *all* tasks are automated. If only a fraction of the tasks are automated, then the scarce factor (labor) will dominate, and growth rates do not explode. We show in this section that with Cobb-Douglas production, a Type II singularity can occur as long as a sufficient fraction of the tasks are automated. In this sense, the singularity might not even require full automation.

Suppose the production function for goods is $Y_t = A_t^{\sigma} K_t^{\alpha} L^{1-\alpha}$ (a constant population simplifies the analysis, but exogenous population growth would not change things). The capital accumulation equation and the idea production function are then specified as

$$\dot{K}_t = \bar{s} L A_t^{\sigma} K_t^{\alpha} - \delta K_t, \tag{34}$$

$$\dot{A}_t = K_t^{\beta} S^{\lambda} A_t^{\phi}, \tag{35}$$

where $0 < \alpha < 1$ and $0 < \beta < 1$, and where we also take S (research effort) to be constant. Following the Zeira (1998) model discussed earlier, we interpret α as the fraction of goods tasks that have been automated and β as the fraction of tasks in idea production that have been automated.

The standard endogenous growth result requires "constant returns to accumulable factors." To see what this means, it is helpful to define a key parameter:

$$\gamma := \gamma \frac{\sigma}{1-\alpha} \cdot \frac{\beta}{1-\phi}. \tag{36}$$

In this setup, the endogenous growth case corresponds to $\gamma = 1$. Not surprisingly, then, the singularity case occurs if $\gamma > 1$. Importantly, notice that this can occur with both α and β less than one, that is, when tasks are not fully automated. For example, in the case in which $\alpha = \beta = \phi = 1/2$, then $\gamma = 2 \cdot \sigma$, so explosive growth and a singularity will occur if $\sigma > 1/2$. We show that $\gamma > 1$ delivers a Type II singularity in the remainder of this section. The argument builds on the argument given in the previous subsection.

In growth rates, the laws of motion for capital and ideas are

$$\frac{\dot{K}_t}{K_t} = \bar{s} L^{1-\alpha} \frac{A_t^{\sigma}}{K_t^{1-\alpha}} - \delta, \tag{37}$$

$$\frac{\dot{A}_t}{A_t} = S^{\lambda} \frac{K_t^{\beta}}{A_t^{1-\phi}}. \tag{38}$$

It is easy to show that these growth rates cannot be constant if $\gamma > 1$.[14]

If the growth rates are rising over time to infinity, then eventually either $g_{At} > g_{Kt}$, or the reverse, or the two growth rates are the same. Consider the first case, that is, $g_{At} > g_{Kt}$; the other cases follow the same logic. Once again, to simplify the algebra, set $\delta = 0$, $S = 1$, and $\bar{s}L^{1-\alpha} = 1$. Multiplying the growth rates together in this case gives

$$\frac{\dot{A}_t}{A_t} \cdot \frac{\dot{K}_t}{K_t} = \frac{K_t^{\beta}}{A_t^{1-\phi}} \cdot \frac{A_t^{\sigma}}{K_t^{1-\alpha}}. \tag{39}$$

Since $g_A > g_K$, we then have

$$\begin{aligned} \left(\frac{\dot{A}_t}{A_t}\right)^2 &> \frac{K_t^{\beta}}{A_t^{1-\phi}} \cdot \frac{A_t^{\sigma}}{K_t^{1-\alpha}} \\ &> \frac{1}{K_t} \cdot \frac{K_t^{\beta}}{A_t^{1-\phi}} \cdot \frac{A_t^{\sigma}}{K_t^{1-\sigma}} && \text{(since } K_t > 1 \text{ eventually)} \\ &> \frac{1}{K_t^{1-\beta}} \cdot \frac{1}{A_t^{1-\phi}} \cdot \frac{A_t^{\sigma}}{K_t^{1-\sigma}} && \text{(rewriting)} \\ &> \frac{1}{A_t^{1-\beta}} \cdot \frac{1}{A_t^{1-\phi}} \cdot \frac{A_t^{\sigma}}{A_t^{1-\alpha}} && \text{(since } A_t > K_t \text{ eventually)} \\ &> A_t^{\gamma-1} && \text{(collecting terms).} \end{aligned}$$

Therefore,

$$\frac{\dot{A}_t}{A_t} > A_t^{(\gamma-1)/2}. \tag{40}$$

With $\gamma > 1$, the growth rate grows at least as fast as A_t raised to a positive power. But even if it grew just this fast we would have a singularity, by the same arguments given before. The case with $g_{Kt} > g_{At}$ can be handled in the same way, using Ks instead of As. QED.

Example 4: Singularities via Superintelligence

The examples of growth explosions above are based in automation. These examples can also be read as creating "superintelligence" as an artifact of automation, in the sense that advances of A_t across all tasks include, implicitly, advances across cognitive tasks, and hence a resulting singularity can be conceived of as commensurate with an intelligence explosion. It is interesting that automation itself can provoke the emergence of superintelligence. However, in the telling of many futurists, the story runs differently, where

[14] If the growth rate of K is constant, then $\sigma g_A = (1 - \alpha)g_K$, so K is proportional to $A^{\sigma/(1-\alpha)}$. Making this substitution in equation (35) and using $\gamma > 1$ then implies that the growth rate of A would explode, and this requires the growth rate of K to explode.

an intelligence explosion occurs first and then, through the insights of this superintelligence, a technological singularity may be reached. Typically the AI is seen as "self-improving" through a recursive process.

This idea can be modeled using similar ideas to those presented above. To do so in a simple manner, divide tasks into two types: physical and cognitive. Define a common level of intelligence across the cognitive tasks by a productivity term $A_{\text{cognitive}}$, and further define a common productivity at physical tasks, A_{physical}. Now imagine we have a unit of AI working to improve itself, where progress follows

$$\dot{A}_{\text{cognitive}} = A_{\text{cognitive}}^{1+\omega}. \tag{41}$$

We have studied this differential equation above, but now we apply it to cognition alone. If $\omega > 0$, then the process of self-improvement explodes, resulting in an unbounded intelligence in finite time.

The next question is how this superintelligence would affect the rest of the economy. Namely, would such superintelligence also produce an output singularity? One route to a singularity could run through the goods production function: to the extent that physical tasks are not essential (i.e., $\rho \geq 0$), then the intelligence explosion will drive a singularity in output. However, it seems noncontroversial to assert that physical tasks are essential to producing output, in which case the singularity will have potentially modest effects directly on the goods production channel.

The second route lies in the idea production function. Here the question is how the superintelligence would advance the productivity at physical tasks, $\text{A}_{\text{physical}}$. For example, if we write

$$\dot{A}_{\text{physical}} = A_{\text{cognitive}}^{\gamma} F(K,L), \tag{42}$$

where $\gamma > 0$, then it is clear that A_{physical} will also explode with the intelligence explosion. That is, we imagine that the superintelligent AI can figure out ways to vastly increase the rate of innovation at physical tasks. In the above specification, the output singularity would then follow directly upon the advent of the superintelligence. Of course, the idea production functions (41) and (42) are particular, and there are reasons to believe they would not be the correct specifications, as we will discuss in the next section.

9.4.2 Objections to Singularities

The above examples show ways in which automation may lead to rapid accelerations of growth, including ever-increasing growth rates or even a singularity. Here we can consider several possible objections to these scenarios, which can broadly be characterized as "bottlenecks" that AI cannot resolve.

Automation Limits

One kind of bottleneck, which has been discussed above, emerges when some essential input(s) to production are not automated. Whether AI can ultimately perform all essential cognitive tasks, or more generally achieve human intelligence, is widely debated. If not, then growth rates may still be larger with more automation and capital intensity (sections 9.2 and 9.3), but the "labor free" singularities featured above (section 9.4.1) become out of reach.

Search Limits

A second kind of bottleneck may occur even with complete automation. This type of bottleneck occurs when the creative search process itself prevents especially rapid producitivy gains. To see this, consider again the idea production function. In the second example above, we allow for complete automation and show that a true mathematical singularity can ensue. But note also that this result depends on the parameter ϕ. In the differential equation

$$\dot{A}_t = A_t^{1+\phi}$$

we will have explosive growth only if $\phi > 0$. If $\phi \le 0$, then the growth rate declines as A_t advances. Many models of growth and associated evidence suggest that, on average, innovation may be becoming harder, which is consistent with low values of ϕ on average.[15] Fishing out or burden of knowledge processes can point toward $\phi < 0$. Interestingly, the burden of knowledge mechanism (Jones 2009), which is based on the limits of human cognition, may not restrain an AI if an AI can comprehend a much greater share of the knowledge stock than a human can. Fishing-out processes, however, viewed as a fundamental feature of the search for new ideas (Kortum 1997), would presumably also apply to an AI seeking new ideas. Put another way, AI may resolve a problem with the fishermen, but it would not change what is in the pond. Of course, fishing-out search problems can apply not only to overall productivity but also to the emergence of a superintelligence, limiting the potential rate of an AI program's self-improvement (see equation [41]), and hence limiting the potential for growth explosions through the superintelligence channel.

Baumol Tasks and Natural Laws

A third kind of bottleneck may occur even with complete automation and even with a superintelligence. This type of bottleneck occurs when an essential input does not see much productivity growth. That is, we have another form of Baumol's cost disease.

To see this, generalize slightly the task-based production function (5) of section 9.2 as

$$Y = \left[\int_0^1 \left(a_{it}X_{it}\right)^{\rho} di\right]^{1/\rho}, \ \rho < 0,$$

[15] See, for example, Jones (1995), Kortum (1997), Jones (2009), Gordon (2016), and Bloom et al. (2017).

where we have introduced task-specific productivity terms, a_{it}.

In contrast to our prior examples, where we considered a common technology term, A_t, that affected all of aggregate production, here we imagine that productivity at some tasks may be different than others and may proceed at different rates. For example, machine computation speeds have increased by a factor of about 10^{11} since World War II.[16] By contrast, power plants have seen modest efficiency gains and face limited prospects given constraints like Carnot's theorem. This distinction is important, because with $\rho < 0$, output and growth end up being determined not by what we are good at, but by what is essential but hard to improve.

In particular, let's imagine that some superintelligence somehow does emerge, but that it can only drive productivity to (effectively) infinity in a share θ of tasks, which we index from $i \in [0,\theta]$. Output thereafter will be

$$Y = \left[\int_{\theta}^{1} (a_{it} Y_{it})^{\rho} \, di \right]^{1/\rho}.$$

Clearly, if these remaining technologies a_{it} cannot be radically improved, we no longer have a mathematical singularity (Type II growth explosion) and may not even have much future growth. We might still end up with an *AK* model, if all the remaining tasks can be automated at low cost, and this can produce at least accelerating growth if the a_{it} can be somewhat improved but, again, in the end we are still held back by the productivity growth in the essential things that we are worst at improving. In fact, Moore's Law, which stands in part behind the rise of artificial intelligence, may be a cautionary tale along these lines. Computation, in the sense of arithmetic operations per second, has improved at mind-boggling rates and is now mind-bogglingly fast. Yet economic growth has not accelerated, and may even be in decline.

Through the lens of essential tasks, the ultimate constraint on growth will then be the capacity for progress at the really hard problems. These constraints may in turn be determined less by the limits of cognition (i.e., traditionally human intelligence limits, which an AI superintelligence may overcome) and more by the limits of natural laws, such as the second law of thermodynamics, which constrain critical processes.[17]

Creative Destruction

Moving away from technological limits per se, the positive effect of AI (and super AI) on productivity growth may be counteracted by another

[16] This ratio compares Beltchley Park's Colossus, the 1943 vacuum tube machine that made 5×10^5 floating point operations per second, with the Sunway TaihuLight computer, which in 2016 peaked at 9×10^{16} operations per second.

[17] Returning to example 4 above, note that equation (42) assumes that all physical constraints can be overcome by superintelligence. However, one might alternatively specify $\max(A_{\text{physical}}) = c$, representing a firm physical constraint.

effect working through creative destruction and its impact on innovation incentives. Thus in the appendix we develop a Schumpeterian model in which: (a) new innovations displace old innovations; and (b) innovations involves two steps, where the first step can be performed by machines but the second step requires human inputs to research. In a singularity-like limit where successive innovations come with no time in between, the private returns to human research and development (R&D) falls down to zero and as a result innovation and growth taper off. More generally, the faster the first step of each successive innovation as a result of AI, the lower the return to human investment in stage-two innovation, which in turn counteracts the direct effect of AI and super-AIon innovation-led growth pointed out above.

9.4.3 Some Additional Thoughts

We conclude this section with additional thoughts on how AI and its potential singularity effects might affect growth and convergence.

A first idea is that new AI technologies might allow imitation/learning of frontier technologies to become automated. That is, machines would figure out in no time how to imitate frontier technologies. Then a main source of divergence might become credit constraints, to the extent that those might prevent poorer countries or regions from acquiring superintelligent machines whereas developed economies could afford such machines. Thus one could imagine a world in which advanced countries concentrate all their research effort on developing new product lines (i.e., on frontier innovation) whereas poorer countries would devote a positive and increasing fraction of their research labor on learning about the new frontier technologies as they cannot afford the corresponding AI devices. Overall, one would expect an increasing degree of divergence worldwide.

A second conjecture is that, anticipating the effect of AI on the scope and speed of imitation, potential innovators may become reluctant to patent their inventions, fearing that the disclosure of new knowledge in the patent would lead to straight imitation. Trade secrets may then become the norm, instead of patenting. Or alternatively innovations would become like what financial innovations are today, that is, knowledge creation with huge network effects and with very little scope for patenting.

Finally, with imitation and learning being performed mainly by supermachines in developed economies, then research labor would become (almost) entirely devoted to product innovation, increasing product variety or inventing new products (new product lines) to replace existing products. Then, more than ever, the decreasing returns to digging deeper into an existing line of product would be offset by the increased potential for discovering new product lines. Overall, ideas might end up being easier to find, if only because of the singularity effect of AI on recombinant idea-based growth.

9.5 Artificial Intelligence, Firms, and Economic Growth

To this point, we have linked artificial intelligence to economic growth emphasizing features of the production functions of goods and ideas. However, the advance of artificial intelligence and its macroeconomic effects will depend on the potentially rich behavior of firms. We have introduced one such view already in the prior section, where considerations of creative destruction provide an incentive-oriented mechanism that may be an important obstacle to singularities. In this section, we consider firms' incentives and behavior more generally to further outline the AI research agenda. We examine potentially first-order issues that emerge when introducing market structure, sectoral differences, and organizational considerations within firms.

9.5.1 Market Structure

Existing work on competition and innovation-led growth points to the existence of two counteracting effects: on the one hand, more intense product market competition (or imitation threat) induces neck-and-neck firms at the technological frontier to innovate in order to escape competition; on the other hand, more intense competition tends to discourage firms behind the current technology frontier to innovate and thereby catch-up with frontier firms. Which of these two effects dominates, in turn, depends upon the degree of competition in the economy, and/or upon how advanced the economy is. While the escape competition effect tends to dominate at low initial levels of competition and in more advanced economies, the discouragement effect may dominate for higher levels of competition or in less advanced economies.[18]

Can AI affect innovation and growth through potential effects it might have on product market competition? A first potential channel is that AI may facilitate the imitation of existing products and technologies. Here we particularly have in mind the idea that AI might facilitate reverse engineering, and thereby facilitate the imitation of leading products and technologies. If we follow the inverted-U logic of Aghion et al. (2005), in sectors with initially low levels of imitation, some AI-induced reverse engineering might stimulate innovation by virtue of the escape-competition effect. But too high (or too immediate) an imitation threat will end up discouraging innovation as potential innovators will face excessive expropriation. A related implication of AI is that its introduction may speed up the process by which each individual sector becomes congested over time. This in turn may translate into faster decreasing returns to innovating within any existing sector (see Bloom et al. 2014), but by the same token it may induce potential innovators to devote more resources to inventing new lines in

[18] For example, see Aghion and Howitt (1992) and Aghion et al. (2005).

order to escape competition and imitation within current lines. The overall effect on aggregate growth will in turn depend upon the relative contributions of within-sector secondary innovation and fundamental innovation aimed at creating new product lines (see Aghion and Howitt 1996) to the overall growth process.

Another channel whereby AI and the digital revolution may affect innovation and growth through affecting the degree of product market competition is in relation to the development of platforms or networks. A main objective of platform owners is to maximize the number of participants to the platform on both sides of the corresponding two-sided markets. For example, Google enjoys a monopoly position as a search platform, Facebook enjoys a similar position as a social network with more than 1.7 billion users worldwide each month, and so does Booking.com for hotel reservations (more than 75 percent of hotel clients resort to this network). And the same goes for Uber in the area of individual transportation, Airbnb for apartment renting, and so on. The development of networks may in turn affect competition in at least two ways. First, data access may act as an entry barrier for creating new competing networks, although it did not prevent Facebook from developing a new network after Google. More important, networks can take advantage of their monopoly positions to impose large fees on market participants (and they do), which may discourage innovation by these participants, whether they are firms or self-employed individuals.

In the end, whether escape competition or discouragement effects dominate will depend upon the type of sector (frontier/neck-and-neck or older/lagging), the extent to which AI facilitates reverse engineering and imitation, and upon competition and/or regulatory policies aimed at protecting intellectual property rights while lowering entry barriers. Recent empirical work (e.g., see Aghion, Howitt, and Prantl 2015) points at patent protection and competition policy being complementary in inducing innovation and productivity growth. It would be interesting to explore how AI affects this complementarity between the two policies.

9.5.2 Sectoral Reallocation

A recent paper by Baslandze (2016) argues that the information technology (IT) revolution has produced a major knowledge diffusion effect, which in turn has induced a major sectoral reallocation from sectors that do not rely much on technological externalities from other fields or sectors (e.g., textile industries) to sectors that rely more heavily on technological externalities from other sectors. Her argument, which we believe applies to AI, rests on the following two counteracting effects of IT on innovation incentives: on the one hand, firms can more easily learn from each other and therefore benefit more from knowledge diffusion from other firms and sectors; on the other hand, the improved access to knowledge from other firms and sectors induced by IT (or AI) increases the scope for business stealing.

In high-tech sectors where firms benefit more from external knowledge, the former effect—knowledge diffusion—will dominate whereas in sectors that do not rely much on external knowledge the latter effect—competition or business stealing—will tend to dominate. Indeed in more knowledge dependent sectors firms see both their productive and their innovative capabilities increase to a larger extent than the capabilities of firms in sectors that rely less on knowledge from other sectors.

It then immediately follows that the diffusion of IT—and AI for our purpose—should lead to an expansion of sectors that rely more on external knowledge (in which the knowledge diffusion effect dominates) at the expense of the more traditional (and more self-contained) sectors where firms do not rely as much on external knowledge.

Thus, in addition to its direct effects on firms' innovation and production capabilities, the introduction of IT and AI involve a knowledge diffusion effect that is augmented by a sectoral reallocation effect at the benefit of high-tech sectors that rely more on knowledge externalities from other fields and sectors. The positive knowledge diffusion effect is partly counteracted by the negative business-stealing effect (Baslandze shows that the latter effect has been large in the United States and that without it the IT revolution would have yet induced a much higher acceleration in productivity growth for the whole US economy).

Based on her analysis, Baslandze (2016) responds to Gordon (2012) with the argument that Gordon only took into account the direct effect of IT and not its indirect knowledge diffusion and sectoral reallocation effects on aggregate productivity growth.

We believe that the same points can be made with respect to AI instead of IT, and one could try and reproduce Baslandze's calibration exercise to assess the relative importance of the direct and indirect effects of AI, to decompose the indirect effect of AI into its positive knowledge diffusion effect and its potentially negative competition effect, and to assess the extent to which AI affects overall productivity growth through its effects on sectoral reallocation.

9.5.3 Organization

How should we expect firms to adapt their internal organization, the skill composition of their workforce and their wage policies to the introduction of AI? In his recent book, *Economics for the Common Good*, Tirole (2017) spells out what one may consider to be "common wisdom" expectations on firms and AI. Namely, introducing AI should: (a) increase the wage gap between skilled and unskilled labor, as the latter is presumably more substitutable to AI than the former; (b) the introduction of AI allows firms to automate and dispense with middle men performing monitoring tasks (in other words, firms should become flatter, that is, with higher spans of control); (c) should encourage self-employment by making it easier for individuals to build their reputation. Let us revisit these various points in more detail. AI, skills, and wage premia: on AI and the increased gap between skilled and unskilled wage, the prediction brings us back to Krusell et al. (2000) based on an aggregate production function in which physical equipment is more substitutable to unskilled labor than to skilled labor, these authors argued that the observed acceleration in the decline of the relative price of production equipment goods since the mid-1970s could account for most of the variation in the college premium over the past twenty-five years. In other words, the rise in the college premium could largely be attributed to an increase in the rate of (capital-embodied) skill-biased technical progress. And, presumably, AI is an extreme form of capital-embodied, skill-biased technical change, as robots substitute for unskilled labor but require skilled labor to be installed and exploited. However, recent work by Aghion et al. (2017) suggests that while the prediction of a premium to skills may hold at the macroeconomic level, it perhaps misses important aspects of firms' internal organization and that the organization itself may evolve as a result of introducing AI. More specifically, Aghion et al. (2017) use matched employer-employee data from the United Kingdom, which they augment with information on R&D expenditures, to analyze the relationship between innovativeness and average wage income across firms.

A first, not surprising, finding is that more R&D-intensive firms pay higher wages on average and employ a higher fraction of high-occupation workers than less R&D-intensive firms (see figure 9.4).

This, in turn, is perfectly in line with the above prediction (a) but also with prediction (b) as it suggests that more innovative (or more "frontier") firms rely more on outsourcing for low-occupation tasks. However, a more surprising finding in Aghion et al. (2017) is that lower-skill (lower occupation) workers benefit more from working in more R&D-intensive firms (relative to working in a firm that does no R&D) than higher-skill workers. This finding is summarized by figure 9.5. In that figure, we first see that higher-skill workers earn more than lower-skill workers in any firm no matter how R&D intensive that firm is (the high-skill wage curve always lies strictly above the middle-skill curve, which itself always lies above the lower-skill curve). But, more interestingly, the lower-skill curve is steeper than the middle-skill and higher-skill curve. But the slope of each of these curves precisely reflects the premium for workers with the corresponding skill level to working in a more innovative firm.

Similarly, we should expect more AI-intensive firms to: (a) employ a higher fraction of (more highly paid) high-skill workers, (b) outsource an increasing fraction of low-occupation tasks, and (c) give a higher premium to those low-occupation workers they keep within the firm (unless we take the extreme view that all the functions to be performed by low-occupation workers could be performed by robots).

To rationalize the above findings and these latter predictions, let us fol-

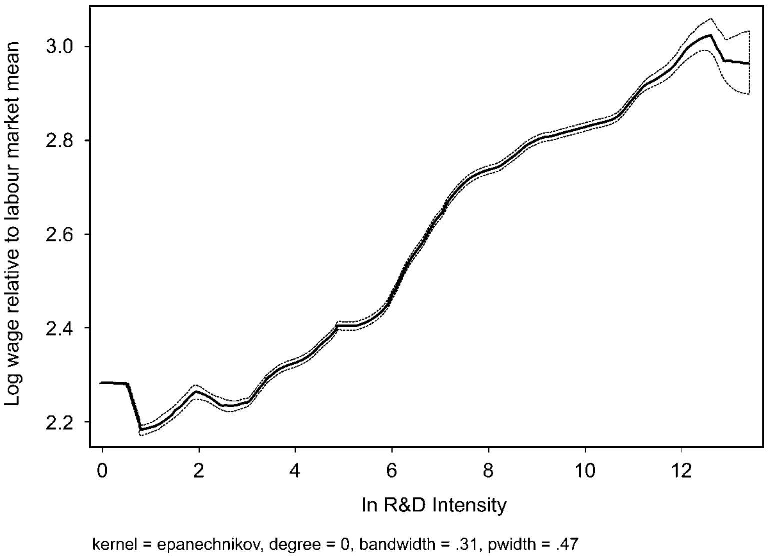

Fig. 9.4 Log hourly wage and R&D intensity

Source: Aghion et al. (2017).

Note: This figure plots the logarithm of total hourly income against the logarithm of total R&D expenditures (intramural + extramural) per employee (R&D intensity).

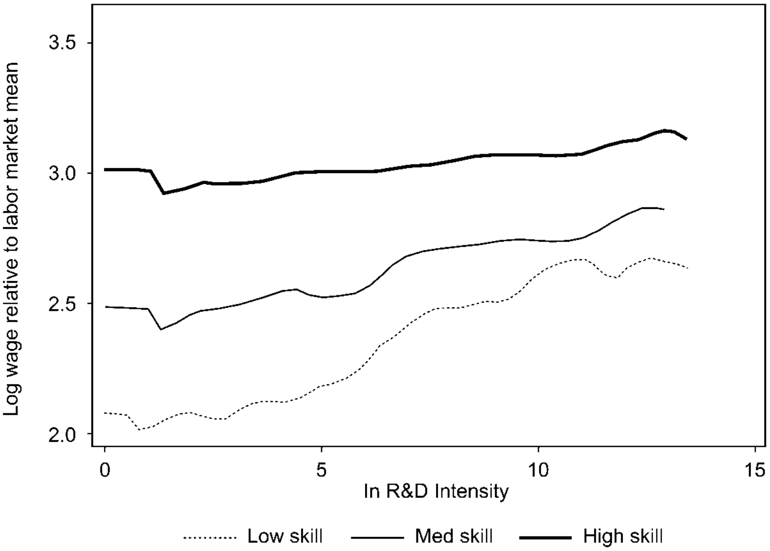

Fig. 9.5 Log hourly wage and R&D intensity

Source: Aghion et al. (2017).

Note: This figure plots the logarithm of total hourly income against the logarithm of total R&D expenditures (intramural + extramural) per employee (R&D intensity) for different skill groups.

low Aghion et al. (2017) who propose a model in which more innovative firms display a higher degree of complementarity between low-skill workers and the other production factors (capital and high-skill labor) within the firm. Another feature of their model is that high-occupation employees' skills are less firm-specific than low-skill workers: namely, if the firm was to replace a high-skill worker by another high-skill worker, the downside risk would be limited by the fact that higher-skill employees are typically more educated employees, whose market value is largely determined by their education and accumulated reputation, whereas low-occupation employees' quality is more firm-specific. This model is meant to capture the idea that low-occupation workers can have a potentially more damaging effect on the firm's value if the firm is more innovative (or more AI intensive for our purpose).

In particular, an important difference with the common wisdom, is that here innovativeness (or AI intensity) impacts on the organizational form of the firm and in particular on complementarity or substitutability between workers with different skill levels within the firm, whereas the common wisdom view takes this complementarity or substitutability as given. Think of a low-occupation employee (e.g., an assistant) who shows outstanding ability, initiative, and trustworthiness. That employee performs a set of tasks for which it might be difficult or too costly to hire a high-skill worker; furthermore, and perhaps more important, the low-occupation employee is expected to stay longer in the firm than higher-skill employees, which in turn encourages the firm to invest more in trust-building and firm-specific human capital and knowledge. Overall, such low-occupation employees can make a big difference to the firm's performance.

This alternative view of AI and firms is consistent with the work of theorists of the firm such as Luis Garicano. Thus in Garicano (2000) downstream, low-occupation employees are consistently facing new problems; among these new problems they sort out are those they can solve themselves (the easier problems) and the more difficult questions they pass on to upstream—higher-skill—employees in the firm's hierarchy. Presumably, the more innovative or more AI-intensive the firm is, the harder it is to solve the more difficult questions, and therefore the more valuable the time of upstream high-occupation employees becomes; this in turn makes it all the more important to employ downstream, low-occupation employees with higher ability to make sure that less problems will be passed on to the upstream, high-occupation employees within the firm so that these high-occupation employees will have more free time to concentrate on solving the most difficult tasks. Another interpretation of the higher complementarity between low-occupation and high-occupation employees in more innovative (or more AI-intensive) firms, is that the potential loss from unreliable low-occupation employees is bigger in such firms: hence the need to select out those low-occupation employees that are not reliable.

This higher complementarity between low-occupation workers and other production factors in more innovative (or more AI-intensive) firms in turn increases the bargaining power of low-occupation workers within the firm (it increases their Shapley Value if we follow Stole and Zwiebel [1996]). This in turn explains the higher payoff for low-occupation workers. It also predicts that job turnover should be lower (tenure should be higher) among low-occupation workers who work for more innovative (more AI-intensive) firms than for low-occupation workers who work for less innovative firms, whereas the turnover difference should be less between high-occupation workers employed by these two types of firms. This additional prediction is also confronted to the data in Aghion et al. (2017).

Note that so far R&D investment has been used as the measure of the firm's innovativeness or frontierness. We would like to test the same predictions, but using explicit measures of AI intensity as the RHS variable in the regressions (investment in robots, reliance on digital platforms). Artificial intelligence and firm organizational form: recent empirical studies (e.g., see Bloom et al. 2014) have shown that the IT revolution has led firms to eliminate middle-range jobs and move toward flatter organizational structure. The development of AI should reinforce that trend, while perhaps also reducing the ratio to low-occupation to high-occupation jobs within firms as we argued above.

A potentially helpful framework to think about firms' organizational forms is Aghion and Tirole (1997). There, a principal can decide whether or not to delegate authority to a downstream agent. She can delegate authority in two ways: (a) by formally allocating control rights to the agent (in that case we say that the principal delegates formal authority to the agent); or (b) informally through the design of the organization, for example, by increasing the span of control or by engaging in multiple activities: these devices enable the principal to commit to leave initiative to the agent (in that case we say that the principal delegates real authority to the agent). And agents' initiative particularly matters if the firm needs to be innovative, which is particularly the case for more frontier firms in their sectors. Whether she decides to delegate formal or only real authority to her agent, the principal faces the following trade-off: more delegation of authority to the agent induces the agent to take more initiative; on the other hand, this implies that the principal will lose some control over the firm, and therefore face the possibility that suboptimal decisions (from her viewpoint) be taken more often. Which of these two counteracting effects of delegation dominates, will in turn depend upon the degree of congruence between the principal's and the agent's preference, but also about the principal's ability to reverse suboptimal decisions.

How should the introduction of AI affect this trade-off between loss of control and initiative? To the extent that AI makes it easier for the principal to monitor the agent, more delegation of authority will be required in order to still elicit initiative from the agent. The incentive to delegate more authority to downstream agents, will also be enhanced by the fact that with AI, suboptimal decision-making by downstream agents can be more easily corrected and reversed: in other words, AI should reduce the loss of control involved in delegating authority downstream. A third reason for why AI may encourage decentralization in decision-making has to do with coordination costs: namely, it may be costly for the principal to delegate decision-making to downstream units if this prevents these units from coordinating within the firm (see Hart and Holmstrom 2010). But here again, AI may help overcome this problem by reducing the monitoring costs between the principal and its multiple downstream units, and thereby induce more decentralization of authority.

More delegation of authority in turn can be achieved through various means: in particular, by eliminating intermediate layers in the firm's hierarchy, by turning downstream units into profit centers or fully independent firms, or through horizontal integration that will commit the principal to spending time on other activities. Overall, one can imagine that the development of AI in more frontier sectors should lead to larger and more horizontally integrated firms, to flatter firms with more profit centers, which outsource an increasing number of tasks to independent self-employed agents. The increased reliance on self-employed independent agents will in turn be facilitated by the fact that, as well explained by Tirole (2017), AI helps agents to quickly develop individual reputations. This brings us to the third aspect of AI and organizations on self-employment. Artificial intelligence and self-employment: as stressed above, AI favors the development of self-employment for at least two reasons: first, it may induce AI intensive firms to outsource tasks, starting with low-occupation tasks; second, it makes it easier for independent agents to develop individual reputations. Does that imply that AI should result in the end of large integrated firms with individuals only interacting with each other through platforms? And which agents are more likely to become self-employed?

On the first question: Tirole (2017) provides at least two reasons for why firms should survive the introduction of AI. First, some activities involve large sunk costs and/or large fixed costs that cannot be borne by a single individual. Second, some activities involve a level of risk-taking that also may not be borne by one single agent. To this we should add the transaction cost argument that vertical integration facilitates relation-specific investments in situations of contractual incompleteness: Can we truly imagine that AI will by itself fully overcome contractual incompleteness?

On the second question: our above discussion suggests that low-skill activities involving limited risk and for which AI helps develop individual reputations (hotel or transport services, health assistance to the elder and/or handicapped, catering services, house cleaning) are primary candidates for increasingly becoming self-employment jobs as AI diffuses in the economy.

And indeed recent studies by Saez (2010), Chetty et al. (2011), and Kleven and Waseem (2013) point to low-income individuals being more responsive to tax or regulatory changes aimed at facilitating self-employment. Natural extensions of these studies would be to explore the extent to which such regulatory changes have had more impact in sectors with higher AI penetration.

The interplay between AI and self-employment also involves potentially interesting dynamic aspects. Thus it might be worth looking at whether self-employment helps individuals accumulate human capital (or at least protects them against the risk of human capital depreciation following the loss of a formal job), and the more so in sectors with higher AI penetration. Also interesting would be to look at how the interplay between self-employment and AI is itself affected by government policies and institutions, and here we have primarily in mind education policy and social or income insurance for the self-employed. How do these policies affect the future performance of currently self-employed individuals, and are they at all complemented by the introduction of AI? In particular, do currently self-employed individuals move back to working for larger firms, and how does the probability of moving back to a regular employment vary with AI, government policy, and the interplay between the two? Presumably, a more performing basic education system and a more comprehensive social insurance system should both encourage self-employed individuals to better take advantage of AI opportunities and support to accumulate skills and reputation and thereby improve their future career prospects. On the other hand, some may argue that AI will have a discouraging effect on self-employed individuals, if it lowers their prospects of ever reintegrating a regular firm in the future, as more AI-intensive firms reduce their demand for low-occupation workers.

9.6 Evidence on Capital Shares and Automation to Date

Models that conceptualize AI as a force of increasing automation suggest that an upswing in automation may be seen in the factor payments going to capital—the capital share. In recent years, the rise in the capital share in the United States and around the world has been a central topic of research. For example, see Karabarbounis and Neiman (2013), Elsby, Hobijn, and Şahin (2013), and Kehrig and Vincent (2017). In this section, we explore this evidence, first for industries within the United States, second for the motor vehicles industry in the United States and Europe, and finally by looking at how changes in capital shares over time correlate with the adoption of robots.

Figure 9.6 reports capital shares by industry from the US KLEMS data of Jorgenson, Ho, and Samuels (forthcoming); shares are smoothed using an HP filter with smoothing parameter 400 to focus on the medium- to long-run trends. It is well-known that the aggregate capital share has increased since at least the year 2000 in the US economy. Figure 9.6 shows that this aggregate trend holds up across a large number of sectors, including agriculture, construction, chemicals, computer equipment manufacturing, motor vehicles, publishing, telecommunications, and wholesale and retail trade. The main place where one does not see this trend is in services, including education, government, and health. In those sectors, the capital share is relatively stable or perhaps increasing slightly since 1990. But the big trend one sees in these data from services is a large downward trend between 1950 and 1980. It would be interesting to know more about what accounts for this trend.

While the facts are broadly consistent with automation (or an increase in automation), it is also clear that capital and labor shares involve many other economic forces as well. For example, Autor et al. (2017) suggest that a composition effect involving a shift toward superstar firms with high capital shares underlies the industry trends. That paper and Barkai (2017) propose that a rise in industry concentration and markups may underlie some of the increases in the capital share. Changes in unionization over time may be another contributing factor to the dynamics of factor shares. This is all to say that a much more careful analysis of factor shares and automation is required before any conclusions can be drawn.

Keeping that important caveat in mind, figure 9.7 shows evidence on the capital share in the manufacturing of transportation equipment for the United States and several European countries. As Acemoglu and Restrepo (2017) note (more on this below), the motor vehicles industry is by far the industry that has invested most heavily in industrial robots during the past two decades, so this industry is particularly interesting from the standpoint of automation.

The capital share in transportation equipment (including motor vehicles, but also aircraft and shipbuilding) shows a large increase in the United States, France, Germany, and Spain in recent decades. Interestingly, Italy and the United Kingdom exhibit declines in this capital share since 1995. The absolute level differences in the capital share for transportation equipment in 2014 are also interesting, ranging from a high of more than 50 percent in the United States to a low of around 20 percent in recent years in the United Kingdom. Clearly it would be valuable to better understand these large differences in levels and trends. Automation is likely only a part of the story.

Acemoglu and Restrepo (2017) use data from the International Federation of Robots to study the impact of the adoption of industrial robots on the US labor market. At the industry level, this data is available for the decade 2004 to 2014. Figure 9.8 shows data on the change in capital share by industry versus the change in the use of industrial robots.

Two main facts stand out from the figure. First, as noted earlier, the motor

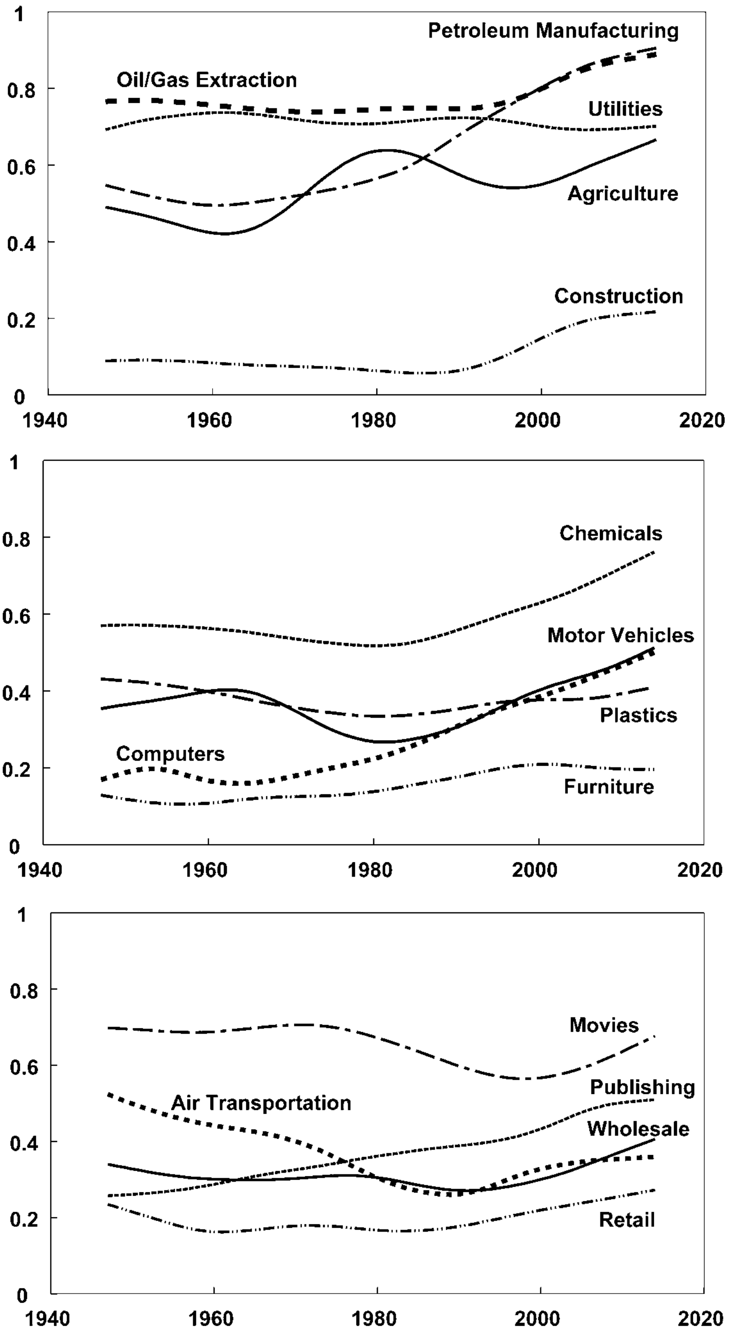

Fig. 9.6 US capital shares by industry

Source: The graph reports capital shares by industry from the U.S. KLEMS data of Jorgenson, Ho, and Samuels (2017).

Note: Shares are smoothed using an HP filter with smoothing parameter 400.

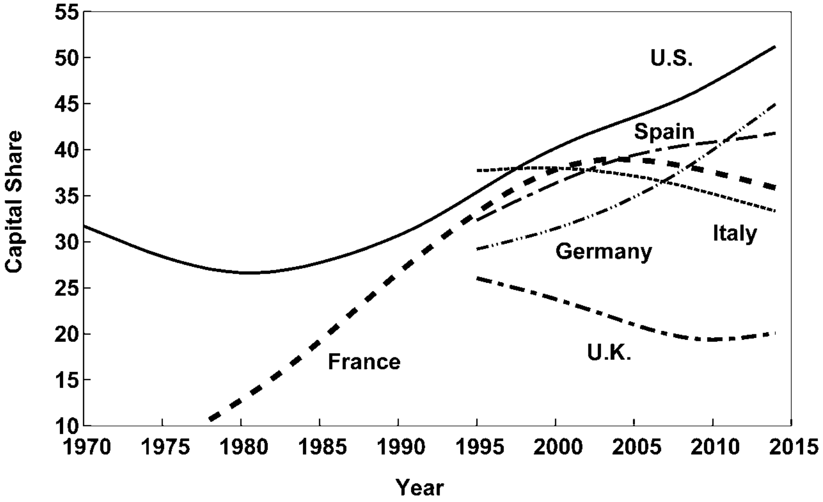

Fig. 9.7 The capital share for transportation equipment

Sources: Data for the European countries are from the EU-KLEMS project (http://www.euklems.net/) for the "transportation equipment" sector, which includes motor vehicles, but also aerospace and shipbuilding; see Jägger (2016). US data are from Jorgenson, Ho, and Samuels (2017) for motor vehicles.

Note: Shares are smoothed using an HP filter with smoothing parameter 400.

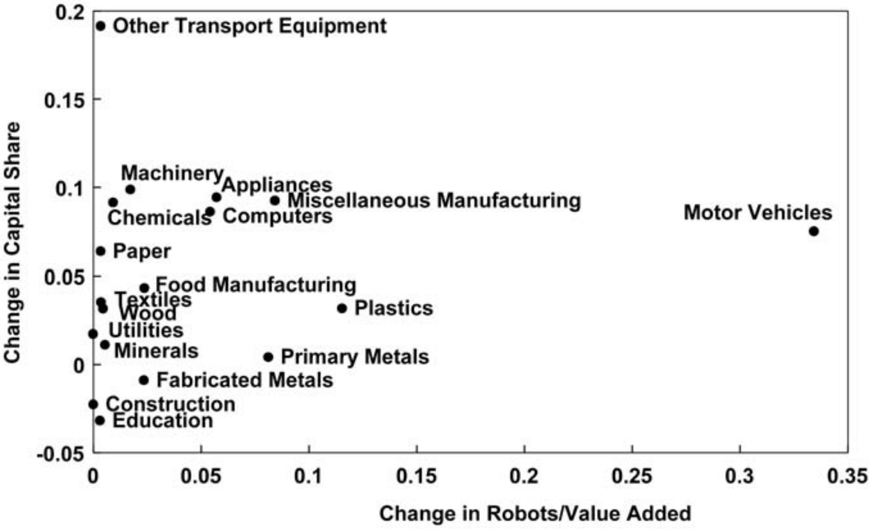

Fig. 9.8 Capital shares and robots, 2004–2014

Sources: The graph plots the change in the capital share from Jorgenson, Ho, and Samuels (2017) against the change in the stock of robots relative to value added using the robots data from Acemoglu and Restrepo (2017).

vehicles industry is by far the largest adopter of industrial robots. For example, more than 56 percent of new industrial robots purchased in 2014 were installed in the motor vehicles industry, the next highest share was under 12 percent in computers and electronic products.

Second, there is little correlation between automation as measured by robots and the change in the capital share between 2004 and 2014. The overall level of industrial robot penetration is relatively small, and as we discussed earlier, other forces including changes in market power, unionization, and composition effects are moving capital shares around in a way that makes it hard for a simple data plot to disentangle.

Graetz and Michaels (2017) conduct a more formal econometric study using the EU-KLEMS data and the International Federation of Robotics data from 1993 until 2007, studying the effect of robot adoption on wages and productivity growth. Similar to what we show in figure 9.8, they find no systematic relationship between robot adoption and factor shares. They do suggest that adoption is associated with boosts to labor productivity.

9.7 Conclusion

In this chapter, we discussed potential implications of AI for the growth process. We began by introducing AI in the production function of goods and services and tried to reconcile evolving automation with the observed stability in the capital share and per capita GDP growth over the last century. Our model, which introduces Baumol's "cost disease" insight into Zeira's model of automation, generates a rich set of possible outcomes. We thus derived sufficient conditions under which one can get overall balanced growth with a constant capital share that stays well below 100 percent, even with nearly complete automation. Essentially, Baumol's cost disease leads to a decline in the share of GDP associated with manufacturing or agriculture (once they are automated), but this is balanced by the increasing fraction of the economy that is automated over time. The labor share remains substantial because of Baumol's insight: growth is determined not by what we are good at but rather by what is essential and yet hard to improve. We also saw how this model can generate a prolonged period with high capital share and relatively low aggregate economic growth while automation keeps pushing ahead.

Next, we speculated on the effects of introducing AI in the production technology for new ideas. Artificial intelligence can potentially increase growth, either temporarily or permanently, depending on precisely how it is introduced. It is possible that ongoing automation can obviate the role of population growth in generating exponential growth as AI increasingly replaces people in generating ideas. Notably, in this chapter, we have taken automation to be exogenous and the incentives for introducing AI in various places clearly can have first-order effects. Exploring the details of endogenous automation and AI in this setup is a crucial direction for further research.

We then discussed the (theoretical) possibility that AI could generate some form of a singularity, perhaps even leading the economy to achieve infinite income in finite time. If the elasticity of substitution in combining tasks is less than one, this seems to require that all tasks be automated. But with Cobb-Douglas production, a singularity could occur even with less than full automation because the nonrivalry of knowledge gives rise to increasing returns. Nevertheless, here too the Baumol theme remains relevant: even if many tasks are automated, growth may remain limited due to areas that remain essential yet are hard to improve. Thus in the appendix we show that if some steps in the innovation process require human R&D, then super AI may end up slowing or even ending growth by exacerbating business-stealing, which in turn discourages human investments in innovation. Such possibilities, as well as other implications of "super-AI" (for example for cross-country convergence and property right protection), remain promising directions for future research.

The chapter next considered how firms may influence, and be influenced by, the advance of artificial intelligence, with further implications for understanding macroeconomic outcomes. We considered diverse issues of market structure, sectoral reallocations, and firms' organizational structure. Among the insights here we see that AI may in part discourage future innovation by speeding up imitation; similarly, rapid creative destruction, by limiting the returns to an innovation, may impose its own limit on the growth process. From an organizational perspective, we also conjectured that while AI should be skill-biased for the economy as a whole, more AI-intensive firms are likely to: (a) outsource a higher fraction of low-occupation tasks to other firms, and (b) pay a higher premium to the low-occupation workers they keep inside the firm.

Finally, we examined sectoral-level evidence regarding the evolution of capital shares in tandem with automation. Consistent with increases in the aggregate capital share, the capital share also appears to be rising in many sectors (especially outside services), which is broadly consistent with an automation story. At the same time, evidence linking these patterns to specific measures of automation at the sectoral level appears weak, and overall there are many economic forces at work in the capital share trends. Developing sharper measures of automation and investigating the role of automation in the capital share dynamics are additional, important avenues for further research.

Appendix

Artificial Intelligence in a Schumpeterian Model with Creative Destruction

In this appendix we describe and model a situation in which superintelligence (or "super-AI") may kill growth because it exacerbates creative destruction and thereby discourages any human investment into R&D. We first lay out a basic version of the Schumpeterian growth model. We then extend the model to introduce AI in the innovation technology.

Basics

Time is continuous and individuals are infinitely lived, there is a mass L of individuals who can decide between working in research or in production. Final output is produced according to

$$y = Ax^{\alpha},$$

where x is the flow of intermediate input and A is a productivity parameter measuring the quality of intermediate input x. Each innovation results in a new technology for producing final output and a new intermediate good to implement the new technology. It augments current productivity by the multiplicative factor $\gamma > 1$: $A_{t+1} = \gamma A_t$. Innovations in turn are the (random) outcome of research, and are assumed to arrive discretely with Poisson rate $\lambda.n$ where n is the current flow of research.

In a steady state the allocation of labor between research and manufacturing remains constant over time, and is determined by the arbitrage equation

$$\omega = \lambda\gamma v, \tag{9A.1}$$

where the LHS of (A) is the productivity-adjusted wage rate $\omega = (w/A)$ which a worker earns by working in the manufacturing sector and $\lambda\gamma v$ is the expected reward from investing one unit flow of labor in research. The productivity-adjusted value v of an innovation is determined by the Bellman equation

$$rv = \tilde{\pi}\left(\omega\right) - \lambda nv,$$

where $\tilde{\pi}(\omega)$ denotes the productivity-adjusted flow of monopoly profits accruing to a successful innovator and where the term $(-\lambda nv)$ corresponds to the capital loss involved in being replaced by a subsequent innovator.

The above arbitrage equation, which can be reexpressed as

$$\omega = \lambda\gamma\frac{\tilde{\pi}(\omega)}{r + \lambda n}, \tag{9A.2}$$

together with the labor market-clearing equation

$$\tilde{x}(\omega) + n = L, \tag{9A.3}$$

where $\tilde{x}(\omega)$ is the manufacturing demand for labor, jointly determine the steady-state amount of research n as a function of the parameters $\lambda, \gamma, L, r, \alpha$.

The average growth rate is equal to the size of each step, lnγ, times the average number of innovations per unit of time, ln that is, $g = \lambda n \ln\gamma$.

A Schumpeterian Model with Artificial Intelligence

As before, there are L workers who can engage either in production of existing intermediate goods or in research aimed at discovering new intermediate goods. Each intermediate good is linked to a particular GPT. We follow Helpman and Trajtenberg (1994) in supposing that before any of the intermediate goods associated with GPT can be used profitably in the final goods sector, some minimal number of them must be available. We lose nothing essential by supposing that this minimal number is one. Once the good has been invented, its discoverer profits from a patent on its exclusive use in production, exactly as in the basic Schumpeterian model reviewed earlier.

Thus the difference between this model and the above basic model is that now the discovery of a new generation of intermediate goods comes in *two* stages. First a new GPT must come, and then the intermediate good must be invented that implements that GPT. Neither can come before the other. You need to see the GPT before knowing what sort of good will implement it, and people need to see the previous GPT in action before anyone can think of a new one. For simplicity we assume that no one directs R&D toward the discovery of a GPT. Instead, the discovery arrives as a serendipitous by-product of the collective experience of using the previous one.

Thus the economy will pass through a sequence of cycles, each having two phases; GPT_i arrives at time T_i. At that time the economy enters phase 1 of the i^{th} cycle. During phase 1, the amount n of labor is devoted to research. Phase 2 begins at time $T_i + \Delta_i$ when this research discovers an intermediate good to implement GPT_i. During Phase 2 all labor is allocated to manufacturing until GPT_{i+1} arrives, at which time the next cycle begins.

A steady-state equilibrium is one in which people choose to do the same amount of research each time the economy is in Phase 1, that is, where n is constant from one GPT to the next. As before, we can solve for the equilibrium value of n using a research-arbitrage equation and a labor market-equilibrium curve. Let ω_j be the wage, and v_j the expected present value of the incumbent intermediate monopolist's future profits, when the economy is in phase j, each divided by the productivity parameter A of the GPT currently in use. In a steady state these productivity-adjusted variables will all be independent of which GPT is currently in use.

Because research is conducted in Phase 1 but pays off when the economy enters into Phase 2 with a productivity parameter raised by the factor γ, the usual arbitrage condition must hold in order for there to be a positive level of research in the economy

$$\omega_1 = \lambda\gamma v_2.$$

Suppose that once we are in Phase 2, the new GPT is delivered by a Poisson process with a constant arrival rate equal to m. Then the value of v_2 is determined by the Bellman equation

$$rv_2 = \tilde{\pi}\left(\omega_2\right) + \mu\left(v_1 - v_2\right).$$

By analogous reasoning, we have

$$rv_1 = \tilde{\pi}\left(\omega_1\right) - \lambda n v_1.$$

Combining the above equations yields the research-arbitrage equation

$$\omega_1 = \lambda\gamma\left[\tilde{\pi}(\omega_2) + \frac{\mu\tilde{\pi}(\omega_1)}{r + \lambda n}\right] / \left[r + \mu\right].$$

Because no one does research in Phase 2, we know that the value of ω_2 is determined independently of research, by the market-clearing condition $L = x(\omega_2)$ Thus we can take this value as given and regard the last equation as determining ω_1 as a function of n The value of n is determined, as usual, by this equation together with the labor-market equation

$$L - n = \tilde{x}\left(\omega_1\right).$$

The average growth rate will be the frequency of innovations times the size lng, for exactly the same reason as in the basic model. The frequency, however, is determined a little differently than before because the economy must pass through *two* phases. An innovation is implemented each time a full cycle is completed. The frequency with which this happens is the inverse of the expected length of a complete cycle. This in turn is just the expected length of Phase 1 plus the expected length of Phase 2:

$$1/\lambda n + 1/\mu = \frac{\mu + \lambda n}{\mu\lambda n}.$$

Thus we have the growth equation

$$g = \ln\gamma\frac{\mu\lambda n}{\mu + \lambda n},$$

where n satisfies

$$f(L - n) = \lambda\gamma\left[f(L) + \frac{\mu\tilde{\pi}\left(f(L - n)\right)}{r + \lambda n}\right] / \left[r + \mu\right]$$

with

$$f(.) = \tilde{x}^{-1}(.)$$

as a decreasing function of its argument.

We are interested in the effect of μ on g and in particular by what happens when $\mu \rightarrow \infty$ as a result of AI in the production of ideas. Obviously, $n \rightarrow 0$ when $\mu \rightarrow \infty$ Thus $E = 1/\lambda n + 1/\mu \rightarrow \infty$ and therefore

$$g = \ln\gamma . \frac{1}{E} \rightarrow 0.$$

In other words, we have described and modeled a situation where superintelligence exacerbates creative destruction to a point that all human investments in to R&D are being deterred and as a result growth tapers off. However, two remarks can be made at this stage:

Remark 1: Here, we have assumed that the second innovation stage requires human research only. If instead AI allowed that stage to also be performed by machines, then AI will no longer taper off and can again become explosive as in our core analysis.

Remark 2: We took automation to be completely exogenous and costless. But suppose instead that it costs money to make μ increase to infinity: then, if creative destruction grows without limit as in our analysis above, the incentive to pay for increasing μ will go down to zero since the complementary human R&D for the stage-two innovation is also going to zero. But this goes against having $\mu \rightarrow \infty$ and therefore against having AI kill the growth process.[19]

References

Acemoglu, Daron, and David Autor. 2011. “Skills, Tasks and Technologies: Implications for Employment and Earnings.” In *Handbook of Labor Economics*, vol. 4, edited by O. Ashenfelter and D. Card, 1043–171. Amsterdam: Elsevier.

Acemoglu, Daron, and Pascual Restrepo. 2016. “The Race between Man and Machine: Implications of Technology for Growth, Factor Shares and Employment.” NBER Working Paper no. 22252, Cambridge, MA.

———. 2017. “Robots and Jobs: Evidence from US Labor Markets.” NBER Working Paper no. 23285, Cambridge, MA.

Aghion, Philippe, and Peter Howitt. 1992. “A Model of Growth through Creative Destruction.” *Econometrica* 60 (2): 323–51.

———. 1996. “Research and Development in the Growth Process.” *Journal of Economic Growth* 1 (1): 49–73.

Aghion, Philippe, Peter Howitt, and Susanne Prantl. 2015. “Patent Rights, Product Market Reforms, and Innovation.” *Journal of Economic Growth* 20 (3): 223–62.

Aghion, Philippe, Antonin Bergeaud, Richard Blundell, and Rachel Griffith. 2017. “The Innovation Premium to Low Skill Jobs.” Unpublished manuscript.

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt.

[19] Of course, one could counterargue that super AI becomes increasingly costless in generating new innovation, in which case μ would again go to infinity and growth would again go down to zero.

2005. "Competition and Innovation: An Inverted-U Relationship." *Quarterly Journal of Economics* 120 (2): 701–28.

Aghion, Philippe, and Jean Tirole. 1997. "Formal and Real Authority in Organizations." *Journal of Political Economy* 105 (1): 1–29.

Agrawal, Ajay, John McHale, and Alex Oettl. 2017. "Artificial Intelligence and Recombinant Growth." Unpublished manuscript, University of Toronto.

Alvarez-Cuadrado, Francisco, Ngo Long, and Markus Poschke. 2017. "Capital-Labor Substitution, Structural Change and Growth." *Theoretical Economics* 12 (3): 1229–66.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2017. "The Fall of the Labor Share and the Rise of Superstar Firms." NBER Working Paper no. 23396, Cambridge, MA.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content Of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics* 118 (4): 1279–333.

Barkai, Simcha. 2017. "Declining Labor and Capital Shares." Unpublished manuscript, University of Chicago.

Baslandze, Salome. 2016. "The Role of the IT Revolution in Knowledge Diffusion, Innovation and Reallocation." Meeting Paper no. 1488, Society for Economic Dynamics.

Baumol, William J. 1967. "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis." *American Economic Review* 57:415–26.

Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb. 2017. "Are Ideas Getting Harder to Find?" Unpublished manuscript, Stanford University.

Bloom, Nicholas, Luis Garicano, Raffaella Sadun, and John Van Reenen. 2014. "The Distinct Effects of Information Technology and Communication Technology on Firm Organization." *Management Science* 60 (12): 2859–85.

Boppart, Timo. 2014. "Structural Change and the Kaldor Facts in a Growth Model with Relative Price Effects and Non???Gorman Preferences." *Econometrica* 82:2167–96.

Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri. 2011. "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records." *Quarterly Journal of Economics* 126 (2): 749–804.

Comin, Diego, Danial Lashkari, and Marti Mestieri. 2015. "Structural Transformations with Long-Run Income and Price Effects." Unpublished manuscript, Dartmouth College.

Danaylov, Nikola. 2012. "17 Definitions of the Technological Singularity." Singularity Weblog. https://www.singularityweblog.com/17-definitions-of-the-technological-singularity/.

Elsby, Michael W. L., Bart Hobijn, and Ayşegül Şahin. 2013. "The Decline of the U.S. Labor Share." *Brookings Papers on Economic Activity* 2013 (2): 1–63.

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of Political Economy* 108 (5): 874–904.

Good, I. J. 1965. "Speculations Concerning the First Ultraintelligent Machine." *Advances in Computers* 6: 31–88.

Gordon, Robert J. 2012. "Is U.S. Economic Growth Over? Faltering Innovation Confronts the Six Headwinds." NBER Working Paper no. 18315, Cambridge, MA.

———. 2016. *The Rise and Fall of American Growth: The US Standard of Living since the Civil War*. Princeton, NJ: Princeton University Press.

Graetz, Georg, and Guy Michaels. 2017. "Robots at Work." Unpublished manuscript, London School of Economics.

Grossman, Gene M., Elhanan Helpman, Ezra Oberfield, and Thomas Sampson.

2017. "Balanced Growth Despite Uzawa." *American Economic Review* 107 (4): 1293–312.

Hart, Oliver, and Bengt Holmstrom. 2010. "A Theory of Firm Scope." *Quarterly Journal of Economics* 125 (2): 483–513.

Helpman, Elhanan, and Manuel Trajtenberg. 1998. "A Time to Sow and a Time to Reap: Growth Based on General Purpose Technologies." In *General Purpose Technologies and Economic Growth*, edited by E. Helpman. Cambridge, MA: MIT Press.

Hemous, David, and Morten Olsen. 2016. "The Rise of the Machines: Automation, Horizontal Innovation and Income Inequality." Unpublished manuscript, University of Zurich.

Herrendorf, Berthold, Richard Rogerson, and Akos Valentinyi. 2014. "Growth and Structural Transformation." In *Handbook of Economic Growth*, vol. 2, 855–941. Amsterdam: Elsevier.

Jägger, Kirsten. 2016. "EU KLEMS Growth and Productivity Accounts 2016 release-Description of Methodology and General Notes." The Conference Board Europe.

Jones, Benjamin F. 2009. "The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder?" *Review of Economic Studies* 76 (1): 283–317.

Jones, Charles I. 1995. "R&D-Based Models of Economic Growth." *Journal of Political Economy* 103 (4): 759–84.

———. 2016. "The Facts of Economic Growth." In *Handbook of Macroeconomics*, vol. 2, 3–69 Amsterdam: Elselvier.

Jorgenson, Dale W., Mun S. Ho, and Jon D. Samuels. Forthcoming. "Educational Attainment and the Revival of U.S. Economic Growth." *Education, Skills, and Technical Change: Implications for Future US GDP Growth*, edited by Charles Hulten and Valerie Ramey. Chicago: University of Chicago Press.

Kaldor, Nicholas. 1961. "Capital Accumulation and Economic Growth." In *The Theory of Capital*, edited by F. A. Lutz and D. C. Hague, 177–222. New York: St. Martins Press.

Karabarbounis, Loukas, and Brent Neiman. 2013. "The Global Decline of the Labor Share." *Quarterly Journal of Economics* 129 (1): 61–103.

Kehrig, Matthias, and Nicolas Vincent. 2017. "Growing Productivity without Growing Wages: The Micro-Level Anatomy of the Aggregate Labor Share Decline." Unpublished manuscript, Duke University.

Kleven, Henrik J., and Mazhar Waseem. 2013. "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan." *Quarterly Journal of Economics* 128 (2): 669–723.

Kongsamut, Piyabha, Sergio Rebelo, and Danyang Xie. 2001. "Beyond Balanced Growth." *Review of Economic Studies* 68 (4): 869–82.

Kortum, Samuel S. 1997. "Research, Patenting, and Technological Change." *Econometrica* 65 (6): 1389–419.

Krusell, Per, Lee E. Ohanian, José-Víctor Ríos-Rull, and Giovanni L. Violante. 2000. "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis." *Econometrica* 68 (5): 1029–53.

Kurzweil, Ray. 2005. *The Singularity is Near*. New York: Penguin.

Legg, Shane, and Marcus Hutter. 2007. "A Collection of Definitions of Intelligence." *Frontiers in Artificial Intelligence and Application* 157 (2007): 17–24.

Manuelli, Rodolfo E., and Ananth Seshadri. 2014. "Frictionless Technology Diffusion: The Case of Tractors." *American Economic Review* 104 (4): 1368–91.

Ngai, L. Rachel, and Christopher A. Pissarides. 2007. "Structural Change in a Multisector Model of Growth." *American Economic Review* 97 (1): 429–43.

Nordhaus, William D. 2015. "Are We Approaching an Economic Singularity? Information Technology and the Future of Economic Growth." NBER Working Paper no. 21547, Cambridge, MA.

Peretto, Pietro F., and John J. Seater. 2013. "Factor-Eliminating Technical Change." *Journal of Monetary Economics* 60 (4): 459–73.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5): S71–102.

Saez, Emmanuel. 2010. "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2 (3): 180–212.

Solomonoff, R. J. 1985. "The Time Scale of Artificial Intelligence: Reflections on Social Effects." *Human Systems Management* 5:149–53.

Stole, Lars, and Jeffrey Zwiebel. 1996. "Organizational Design and Technology Choice under Intrafirm Bargaining." 86 (1): 195–222.

Tirole, Jean. 2017. *Economics for the Common Good.* Princeton, NJ: Princeton University Press.

Vinge, Vernor. 1993. "The Coming Technological Singularity: How to Survive in the Post-Human Era." In *Vision-21: Interdisciplinary Science and Engineering in the Era of Cyberspace*, 11–22. Proceedings of a Symposium Coauthored by the NASA Lewis Research Center and the Ohio Aerospace Institute Held in Westlake, Ohio, Mar. 30–31.

Webb, Michael, Greg Thornton, Sean Legassick, and Mustafa Suleyman. 2017. "What Does Artificial Intelligence Do?" Unpublished manuscript, Stanford University.

Weitzman, Martin L. 1998. "Recombinant Growth." *Quarterly Journal of Economics* 113:331–60.

Yudkowsky, Eliezer. 2013. "Intelligence Explosion Microeconomics." Technical Report no. 2013–1, Machine Intelligence Research Institution.

Zeira, Joseph. 1998. "Workers, Machines, and Economic Growth." *Quarterly Journal of Economics* 113 (4): 1091–117.

Comment Patrick Francois

The political economy of artificial intelligence (AI) was not included as a topic in this conference, but political economy arose in a number of conversations, including my discussion of this immensely thought-provoking chapter. So I want to discuss it further here. It is important for two reasons. One, if the scientists' predictions pan out, we are on the cusp of a world where humans will be largely redundant as an economic input. How we manage the relationship between the haves (who own the key inputs) and the have-nots (who only own labor) is going to be a key aspect of societal health. Successful ones will be inclusive in the sense of sharing rents owned by the haves with the have-nots. This is quite obvious. Less obviously, I am going to argue that

Patrick Francois is a professor at the Vancouver School of Economics of the University of British Columbia and a senior fellow at the Canadian Institute for Advanced Research.

For acknowledgments, sources of research support, and disclosure of the author's material financial relationships, if any, please see http://www.nber.org/chapters/c14028.ack.

managing the relationship between high-level human decision-making and our machines servants will involve humans at many levels, no matter how productive machines become. So, even in the limit where machines become better at doing *all* human production, there will still be work for humans in what could be broadly referred to as the political realm.

The chapter of Philippe Aghion, Benjamin Jones, and Charles Jones is a great starting point for the less structured discussion that I am about to set off on here. The chapter explores the growth implications of AI, where the aspect focused on is the increasing automation of production. That is, machines replacing labor at a continually increasing range of production, service, and creative tasks. Automation in this form is not new and has been going on since at least the Industrial Revolution. So any model written down projecting what will/might happen should not run afoul of the basic Kaldor facts. Accordingly, they build a model able to deliver a relatively stable labor share despite the continual displacement of labor from an increasing number of sectors.

In a nutshell this works as follows: with multiple sectors and low enough substitutability across the goods produced in them, consumers spend progressively more of real wealth on sectors not subject to automation. This leads to a protracted relative price increase of nonautomated goods' sectors. So two counteracting forces generate a force toward relative stability of the labor share in their model: (a), labor is usefully employed in fewer sectors—lowering its factor share; but (b), in the sectors where labor continues to work, relative prices are increasing—tending to raise the factor share. Essentially, though progressively fewer things remain useful for humans to do, these things become relatively well remunerated, and this can continue provided there remain *some* things that humans can do better than machines.

But it is when we turn to thinking about what are the products or services where humans will remain essential in production that we start to run into problems. What if humans cannot do anything better than machines? Many discussions at the conference centered around this very possibility. And I must admit that I found the scientists' views compelling on this. Though it has been the case that new services, which have been relatively labor intensive, have emerged as technology has mechanized the production of goods and services, and this has been demonstrated by others (Acemoglu and Restrepo 2016) to be another force that could stabilize the labor share. Even with this, the complete displacement of labor from production of goods and service will arise if machines dominate humans in the performance of *all* tasks.

Scientists disagree on how imminent this eventuality is, but few doubt that it will eventually occur. Though it may well be a limiting case reached only many generations down the track, from now on I will try to imagine what will happen in that limiting case. The one where machines can do everything better than humans. The point I wish to make is that even in such a world where machines are better at all tasks, there will still be an important role for human "work." And that work will involve what will become the almost political task of managing the machines.

The Political Economic Challenges That Machine-Superior Societies Will Face

But before I turn to that, a first challenge societies will face in a completely machine-superior world is: Who owns the machines? Capitalist societies succeed when they create incentives for investment. They reward innovators who come up with and implement good ideas, and thus encourage those ideas. Societies with the features that are well suited to pioneering the advance of machines today are also the economically successful societies, and generally the most healthy societies socially. Incentives for technological advance are greatest where property rights are best protected, and where the taxes on the successful are the lowest. So we predictably see the vanguard of this new world of machine superiority emerging from the most successful capitalist economies like the United States of America.

But everything changes when the machines reach the point of displacing human inputs in the task of innovation, what Aghion, Jones, and Jones term "AI in the idea production function." Here I'm again talking about the extreme case where machines do all of their own innovation much better than people, and without requiring any human input. At this point, the decisions on how to best improve the current technology, the risks to take, the directions to follow, and the implementation are all done by machines. Machines then improve themselves and enter in to a process of creating new and better machines without the need for human intervention.

Aghion, Jones, and Jones developed a fantastically interesting analysis of the almost science fiction-like possibility of singularities and productive extremes that can arise in that stage. I am going to, alternatively, focus on the political economic implications.

Presumably, at least at the start of this period, the human owners of these machines made improvements (and the stream of rents that those improvements generate)that are well identifiable. These are the owners of the machines that did the previous round of inventing. Similarly, as the next generation of improvements emerge, the machines that were earlier invented by the previous machines can be traced back to a primal machine inventor(s) with well-identified human inventor/owners, and so on. In a sense then, this last generation of human inventor/owners will have a claim to the rents generated by the machines from then on.

Should we, as a society, recognize that claim? The answer to that depends on where individuals, the political elites, and the economic elites in that society stand on the issue of inviolability of private property. At the point where machines become self-inventing, redistributing the ownership rents to all individuals in society will come without cost in terms of future growth because human incentives no longer play a role. This won't be easy for many of today's successful societies to do.

The social cost of not doing this will be human unrest on a massive scale. The degree of inequality in a society where the owners of the machines are the last generation of human/inventor/investors and the rest of society earns their incomes from labor will be extreme. Nationalizing ownership of the machines will be costless in terms of future growth, but the elite who own the machines may be (and if history is any guide, will be) extremely reluctant to give up their "hard-earned" rents, and their power, to the passive majority who did not have the foresight, hard work, and luck, to come up with these machines. The societies that will be most functional in this future will be those most willing to tax this last generation of productive inventor/investors to support the unlucky, less able, and perhaps even willingly slothful, who do not own a machine. Countries that, for the very reason of not heavily taxing innovation today will be in the vanguard of creating our technofuture, may have social values that will tend to make them somewhat poorly placed to manage it.

If the elite of such countries succeed in managing to control the political channels whereby rival elites may come to threaten them, or where the excluded masses who do not share ownership of the machines would be able to coordinate against them, they will be able to enjoy machine rents and become almost infinitely richer than the excluded. The autocratic elites of the Soviet Union employed just such methods of exclusion and disruption to rule their countries many decades after they had lost the cooperation of their masses. And they did not have super-smart robots to help them. If the future elite of countries that are willing to protect their rents from owning the economy's productive assets (machines) study history's successful autocrats well enough (or their machines do), this could go on for quite a while.

In contrast, where the machines are nationally owned, and where the rents are shared by all society's members, what I will call inclusive societies, there is no reason that we cannot have equality in consumption. The very good, incentive-based reasons for inequality to exist under capitalism will no longer apply.

The Political Economic Source of Future Human Work

What will humans do for work in a world where machines are better at doing everything than humans? It would seem that the obvious answer is nothing. We will have to learn to create meaning from non-work-related activities, and hopefully overcome our evolved proclivity toward equating personal value with social productivity. I am going to argue that this obvious answer is wrong. There will actually be vital and important work for humans to do in this world, and that the amount of it to be done will be greatest in the most inclusive societies.

Managing the Machines Will Be the Source of Human Work

Why would machines need managing? The machines will be self-replicating, self-maintaining, self-creating, self-repairing, self-improving, so what else needs to be done? What is not so clear is which ends the machines are pursuing.

Usually we tend to think in terms of well-defined human objectives, and for most of these it is a nonquestion as to what machines should do. For example, oncology machines will read MRIs, diagnose potential cancers, order more tests, or operations, or drugs, and so forth, based on protocols they have learned by being run millions of times on training data. They can learn what to do because objectives here are relatively simple, and success in meeting them can be used to determine optimal actions easily. So these machines with very narrow objectives need relatively little managing.

But machines will be producing all output and services in our economy, and while doing this will all the while continually reinvent and modify themselves in pursuit of objectives that were programmed in to them by their human masters. So we will have a complex set of evolving machines who are not only running all production, but doing all inventing as well. We could think of these machines as designed, but through the process of machine learning and machine-based innovation the designs would become far removed from anything imagined by the last generation of human designers that worked on them. Even understanding what they are doing will be difficult for us humans. Perhaps we will develop intuitions about them, a richer human language, or narratives about what they do that will give us some vague understandings of what they are about, but it is reasonable to suppose that no human will fully understand them.

The question is, Will we be willing to let this design direction simply continue without human interjection? I would argue that we will not. We (our societal "we") will be greatly concerned about the direction that this design takes, and managing this direction will require immense human oversight. The more so, the more inclusive a society is. But why would we need to manage it if we have already programmed in to these advanced machines a set of objectives that are human centred? If we have already delegated that to the machines? I am assuming that, as part of this programming, we will find fail-safes to short-circuit rogue machines following objectives that do not advance human welfare, as interestingly sketched by Nick Bostrom (2014), so I am explicitly excluding that particular dystopia.

But even with such fail-safes, additional human involvement will be required. This is because we cannot delegate a particular objective function to machines and be done with it, because whatever delegation that we implement at time t, based on an objective articulated with the knowledge we have at time t, may well be outdated by time $t' > t$ because either our knowledge or our values have changed by t'. We will need people (obviously greatly aided by machines) charged with working out what our social consensus is at time t', informing other citizens at t' what relevant information they need to make their decisions then, and then implementing those changes at time t'. These actions, which would of course be simple for machines to do since they will be so much smarter than us, will be inherently nonimplementable by the machines that are doing all our inventing and production at time t', because those machines will have been programmed with the objective functions of time t society, which is precisely what we wish to countenance changing at time t'.

The whole problem is that writing objectives at time t may lead machines to evolve capacities based on those objectives that become outdated at t'. In order for us to know whether they are outdated at t', we have to first develop a conception of what the machines should be doing at t', and how that differs from what we thought at t, and we need to somehow have a sense of what the machines are actually doing at t' and how it differs from t. All of these things are collective human decisions, and will require immense human effort.

For example, suppose we program in to these advanced machines an objective of maximizing human welfare defined in a utilitarian way in the year 2035. The designing machines will then set off to come up with machine improvements that advance our utilitarian human objectives. But in doing so, they may end up doing some violence to other objectives which, on the whole we were ready as a society to subordinate to sound utilitarian ones in 2035, but are no longer willing to countenance in 2050. For instance, it may be the case that the utilitarian-based inventing machines put no weight on animal welfare, other than how it indirectly advances the utilitarian goal. But it could be that our societal objectives, beliefs, views and so forth have evolved in the intervening years. Maybe we come to learn something more about animal neurology, or maybe we just change our values as we become richer. And then people, on the whole, start to want to privilege other mammals as much as ourselves. Or alternatively perhaps we become so impressed with the complexity of machines that we want to countenance nonorganic life as of value in itself. In either such case, we will need to, as human decision makers, understand enough of what machines are doing in pursuit of some of our earlier objectives to be able to see whether the societal objectives unstated in 2035 are being trammelled upon or not in 2050. They may not be, and in that case nothing much needs to change. But how will we know without checking?

That will be very complicated to do. It firstly requires some humans trying to understand just what it is that the machines are doing in 2050: How they are evolving and what they have been up to? We then need to work out what the relevant parts of that information are for our societal decision makers to know, and in inclusive societies "societal decision makers" are a lot of people. We then need to find a way of communicating this perhaps highly sophisticated information to these decisions makers, some, and perhaps many, of whom have very little technical training about machine function, so that they can make their decisions based on the knowledge and training that they do have.

This process also, of course, begs the question as to who "we" as a set of societal decision makers are in this context, and what "we" want. Some humans must be involved in making these ethical and social decisions. And here I do not mean decisions of the form whether a car should collide with and kill three old citizens instead of a pregnant mother, which is of course difficult, but which we at least implicitly grapple with every day. But I mean the more basic decisions as to what is the societal objective that the network of machines that are not only producing everything for us, but also designing and inventing everything for us are trying to attain. One could argue that we also implicitly engage in such decisions today as a society, for example, when we elect politicians or parties with competing platforms. However, in the future it will be much more explicit, as our collective stance on these things will be needed to determine precisely what direction we will orient our machine inventors to head towards every single day.

It will not be possible (or prudent if it were possible) to delegate this set of conversations and tasks to machines alone. Even though they may be demonstrably smarter and hence better at making those decisions given a well-defined objective function, the point is that there is and never will be such a well-defined social objective function (we have known this since Arrow's impossibility theorem). We need to modify it via our political processes in a continual way, and the objective function followed by the machines will need to be adjusted in reflection of a social conversation that occurs amongst humans. In inclusive societies, where presumably all citizens will have a voice in those decisions, this will involve a lot of people, all of whom will have to be informed so that they can weigh in on that social consensus.

Managing that conversation, reporting back to "us" what is relevant for that conversation emerging from the self-directed world of machines, and then adjusting the trajectory of the machines in light of what "we" decide via whatever social mechanisms we come up with to express as our collective will, must require humans at certain critical points. Human decision making will not be replicable or replaceable by machines here almost by definition.

So, to summarize, I am describing a world that we are admittedly far from today. A world in which most human labor is involved in the set of essentially political tasks related to managing the machines that will be doing all the production in our economy, and hence determining much of our societies' directions. A set of people will need to work at determining just what our current machines are doing and making that intelligible to social decision makers (which in inclusive societies will be a lot of citizens). Another set of people will need to work out how the diverse sets of opinions manifested by citizens maps back to a consensus about what our machines should be doing, and what directions they should be heading toward. All of these workers will be helped by machines, but the machines helping them will need human guidance since they will not be using objective protocols that could ever be unchanging. This is because it is the very protocols that the machines are using that we humans must be constantly discussing changing. Humans, though immeasurably dumber than machines, will be essential and nonsubstitutable in that process.

References

Acemoglu, Daron, and Pascual Restrepo. 2016. "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares and Employment." Unpublished manuscript, Massachusetts Institute of Technology.

Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.