NBER WORKING PAPER SERIES

ARTIFICIAL INTELLIGENCE AND ITS IMPLICATIONS FOR INCOME DISTRIBUTION AND UNEMPLOYMENT

Anton Korinek
Joseph E. Stiglitz

Working Paper 24174
http://www.nber.org/papers/w24174

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2017

We would like to thank our discussant Tyler Cowan as well as Jayant Ray and participants at the NBER conference for helpful comments. We also acknowledge research assistance from Haaris Mateen as well as financial support from the Institute for New Economic Thinking (INET) and the Rewriting the Rules project at the Roosevelt Institute, supported by the Ford and Open Society Foundations, and the Bernard and Irene Schwartz Foundation. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Artificial Intelligence and Its Implications for Income Distribution and Unemployment
Anton Korinek and Joseph E. Stiglitz
NBER Working Paper No. 24174
December 2017
JEL No. D63,E64,O3

ABSTRACT

Inequality is one of the main challenges posed by the proliferation of artificial intelligence (AI) and other forms of worker-replacing technological progress. This paper provides a taxonomy of the associated economic issues: First, we discuss the general conditions under which new technologies such as AI may lead to a Pareto improvement. Secondly, we delineate the two main channels through which inequality is affected – the surplus arising to innovators and redistributions arising from factor price changes. Third, we provide several simple economic models to describe how policy can counter these effects, even in the case of a "singularity" where machines come to dominate human labor. Under plausible conditions, non-distortionary taxation can be levied to compensate those who otherwise might lose. Fourth, we describe the two main channels through which technological progress may lead to technological unemployment – via efficiency wage effects and as a transitional phenomenon. Lastly, we speculate on how technologies to create super-human levels of intelligence may affect inequality and on how to save humanity from the Malthusian destiny that may ensue.

Anton Korinek
Department of Economics
Johns Hopkins University
Wyman Park Building 531
3400 N. Charles Street
Baltimore, MD 21218
and NBER
akorinek@jhu.edu

Joseph E. Stiglitz
Uris Hall, Columbia University
3022 Broadway, Room 212
New York, NY 10027
and NBER
jes322@columbia.edu

1. Introduction

The introduction of artificial intelligence (AI) is the continuation of a long process of automation. Advances in mechanization in the late-nineteenth and early-twentieth century automated much of the physical labor performed by humans. Advances in information technology in the mid- to late-twentieth century automated much of the standardized data processing that used to be performed by humans. However, each of these past episodes of automation left large areas of work that could only be performed by humans.

Some propose that advances in AI are merely the latest wave in this long process of automation (see e.g. Gordon, 2016). Others, by contrast, emphasize that AI critically differs from past inventions: as artificial intelligence draws closer and closer to human general intelligence, much of human labor runs the risk of becoming obsolete and being replaced by AI in all domains. In this view, progress in artificial intelligence is not only a continuation but the culmination of technological progress – it could lead to a course of history that is markedly different from the implications of previous waves of innovation, and may even represent what James Barrat (2013) has termed "Our Final Invention."

No matter what the long-run implications of AI are, it is clear that it has the potential to disrupt labor markets in a major way, even in the short and medium run, affecting workers across many professions and skill levels.[2] The magnitude of these disruptions will depend on two important factors: the speed and the factor bias of progress in AI.

On the first factor, measured productivity has increased rather slowly in recent years, even as the world seems to be captured by AI fever.[3] If AI-related innovations enter the economy at the same slow pace as suggested by recent productivity statistics, then the transition will be slower than e.g. the wave of mechanization in the 1950 – 1970s, and the resulting disruptions may not be very significant. However, there are three possible alternatives: First, some suggest that productivity is significantly under-measured, for example because quality improvements are not accurately captured. The best available estimates suggest that this problem is limited to a few tenth of a percentage point (see e.g. the discussion in Groshen et al., 2017). Furthermore, there are also unmeasured deteriorations in productivity, e.g. declines in service quality as customer service is increasingly automated. Secondly, the aggregate implications of progress in AI may follow a delayed pattern, similar to what happened after the introduction of computers

[2] For example, Frey and Osborne (2017) warn that 47% of jobs in the US economy are at risk of being automated by advances in AI-related fields. Areas in which human intelligence has recently become inferior to artificial intelligence include many applications of radiology, trading in financial markets, paralegal work, underwriting, driving etc.

[3] For example, Google Trends reveals that search interest in the topic "artificial intelligence" has quadrupled over the past four years.

in the 1980s. Robert Solow (1987) famously quipped that "you can see the computer age everywhere but in the productivity statistics." It was not until the 1990s that a significant rise in aggregate productivity could be detected, after sustained investment in computers and a reorganization of business practices had taken place. Third, it is of course possible that a significant discontinuity in productivity growth occurs, as suggested e.g. by proponents of a technological singularity (see e.g. Kurzweil, 2005).

On the second factor, the disruptions generated by AI-related innovations depend on whether they are labor-augmenting or labor-saving, using the terminology of Hicks (1932), i.e. whether at a given wage, the innovations lead to more or less demand for labor. Some suggest that artificial intelligence will mainly *assist* humans in being more productive, and refer to such new technologies as *intelligence assisting innovation, IA,* rather than AI. Although we agree that most AI-related innovations are likely to be complementary to at least some jobs – e.g. the ones applying AI to solve problems – we believe that taking a broader perspective, progress in AI is more likely to substitute for human labor, or even to replace workers outright, as we will assume in some of our formal models below.

We believe that the primary economic challenge posed by the proliferation of AI will be one of income distribution. We economists set ourselves too easy a goal if we just say that technological progress *can make everybody better off* – we also have to say *how we can make this happen*. The following paper is an attempt to do so by discussing some of the key economic research issues that this brings up.[4]

In section 2 of our paper, we provide a general taxonomy of the relationship between technological progress and welfare. We first observe that in a truly first-best economy – in which complete risk markets are available before a veil of ignorance about innovations is lifted – all individuals will share in the benefits of technological progress. However, since the real world does not correspond to this ideal, redistribution is generally needed to ensure that technological progress generates Pareto improvements. If markets are perfect and redistribution is costless, it can always be ensured that technological progress makes everybody better off. The same result holds if the costs of redistribution are sufficiently low. In all these cases, there can be political unanimity about the desirability of technological progress. However, if redistribution is too costly, it may be impossible to compensate the losers of technological progress, and they will rationally oppose progress. Even worse, if the economy suffers from market imperfections, technological progress may actually move the Pareto frontier inwards, i.e. some individuals may necessarily be worse off. Finally, we observe that the

[4] An important, and maybe even more difficult, complementary question, which is beyond the scope of this paper, is to analyze the political issues involved.

first welfare theorem does not apply to the process of innovation, and as a result, privately optimal innovation choices may move the Pareto frontier inwards.

In section 3, we decompose the mechanisms through which innovation leads to inequality into two channels. First, inequality rises because innovators earn a surplus. Unless markets for innovation are fully contestable, the surplus earned by innovators is generally in excess of the costs of innovation and includes what we call innovator rents. We discuss policies that affect the sharing of such rents, such as antitrust policies and changes in intellectual property rights. The second channel is that innovations change the demand for factors such as different types of labor and capital, which affects their prices and generates redistributions. For example, AI may reduce a wide range of human wages and generate a redistribution to entrepreneurs. From the perspective of our first-best benchmark with complete insurance markets, these factor price change represent pecuniary externalities. We discuss policies to counter the effects of the resulting factor price changes.

In section 4, we develop a simple formal model of worker-replacing technological change, i.e. we introduce a machine technology that acts as a perfect substitute for human labor. We study the implications for wages and discuss policy remedies. In the short run, an additional unit of machine labor that is added to the economy earns its marginal product, but also generates a zero-sum redistribution from labor to traditional capital because it changes the relative supply of the two. In the long run, the machine technology turns labor into a reproducible factor. Thus, in the long run, growth will likely be limited by some other irreproducible factor, and all the benefits of technological progress will accrue to that factor. However, since it is in fixed supply, they can be taxed and redistributed without creating distortions, and a Pareto improvement is easily achieved.

In a second model, we demonstrate how changes in patent length and capital taxation can act as a second-best device to redistribute if lump sum transfers between workers and innovators are not available. A longer patent life both delays how quickly innovations enter the public domain, lowering consumer prices, and increases the incentives of innovators to produce worker-replacing machines. However, the resulting losses for workers can be made up for by imposing a distortionary tax on capital and providing transfers, so long as the supply elasticity of capital is sufficiently low. We also discuss the implications of endogenous factor bias in technological change. Worker-replacing technological progress should make capital-saving innovations more desirable, providing some relief to workers. We also note that our economy is developing more and more into a service economy, and that the large role of government in many service sectors (e.g. education, healthcare, etc.) creates ample scope for interventions to support workers.

In section 5, we observe two sound economic reasons that may lead to technological unemployment. The first category of reasons arises because wages cannot adjust, even in the long run: efficiency wage theory implies that employers may find it efficient to pay "fair wages" above the market clearing level so that workers have incentives to exert proper effort. If technological progress continues unabated and the marginal product of workers declines below their cost of living, then classic nutritional efficiency wage theories apply: unemployment would result because workers could not survive working for the market-clearing wage without government support. The second category of technological unemployment arises as a transition phenomenon, when jobs are replaced at a faster rate than workers can find new ones. We discuss a variety of factors that may slow down the adjustment process. Efficiency wage arguments may also play an important role as a transitional phenomenon, in particular if workers' notion of fair wages is sticky. Finally, we discuss that jobs may not only provide wages but also meaning and note that, unless societal attitudes change with the proliferation of AI, it may be welfare enhancing to subsidize jobs rather than simply redistributing resources.

In section 6, we take a longer-term perspective that is somewhat more speculative and discuss the potential implications of super-human artificial intelligence. We consider two scenarios: one in which some humans use technology to enhance themselves and attain super-human intelligence; and one in which autonomous machines that are completely separate from humans reach super-human intelligence. In both cases, the superior productivity of superior intelligence will likely lead to vast increases in income inequality. From a Malthusian perspective, the super-intelligent entities are likely to command a growing share of the scarce resources in the economy, pushing regular humans below their subsistence level. We discuss a number of corrective actions that could be taken.

2. Technological Progress and Welfare: A Taxonomy

In 1930, Keynes wrote an essay on the "Economic Possibilities of our Grandchildren," in which he described how technological possibilities may translate into utility possibilities. He worried about the quality of life that would emerge in a world with excess leisure. And he thought all individuals might face that quandary. But what has happened in recent years has raised another possibility: innovation could lead to a few very rich individuals—who may face this challenge—whereas the vast majority of ordinary workers may be left behind, with wages far below what they were at the peak of the industrial age.

So let us start by considering the arrival of a new technology that partially (or fully) replaces workers and let's ask the question: *would their standard of living necessarily decline*? We will consider this question in a number of different settings, providing a taxonomy for how

technological progress might affect the welfare of different groups in society depending on the environment:

2.1. First Best

We start with a first-best scenario in which we assume that all markets are perfect, including risk markets that allow individuals to insure against the advent of innovations "behind the veil of ignorance," i.e. before they know whether they will be workers or innovators. The main purpose for considering this idealized setting is to demonstrate that from an ex-ante perspective, compensating workers for the losses imposed by technological progress is a question of economic efficiency not redistribution.

If risk markets were perfect and accessible to all agents before they knew their place in the economy, then all agents would be insured against any risk that might significantly affect their well-being, including the risk of innovation reducing the value of their factor endowment. For example, workers would be insured against the risk of declining wages.[5] This leads us to the following observation:

Observation 1) ***Consider a first-best world in which all individuals have access to a perfect insurance market "behind the veil of ignorance," i.e. before they know whether they will be innovators or workers. If an innovation occurs in such a world, the winners would compensate the losers as a matter of optimal risk sharing. As a result, technological progress always makes everybody better off, and there is political unanimity in supporting it.***

This is a powerful observation because it reminds us that if we had an ideal market, something that very much looks like redistribution would naturally emerge. In our first-best economy, there are no losers from technological progress. Losers only exist if risk markets are imperfect compared to this benchmark. In more technical language, worker-replacing technological progress imposes pecuniary externalities on workers, which lead to inefficiency when risk markets are imperfect (see e.g. Stiglitz, 1981; Greenwald and Stiglitz, 1986; Geanakoplos and Polemarchakis, 1986; or more recently Davila and Korinek, 2017).

This implies that policy measures to mitigate or undo the pecuniary externalities arising from technological progress – for example redistribution programs –make the economy's allocation more efficient from an ex-ante perspective, rather than "interfering" with economic efficiency. They bring us closer to the allocation that a well-functioning risk market would achieve. Policymakers who oppose redistribution that compensates the losers of innovation because it interferes with the free market seem to – inappropriately, in our view – take an ex-post

[5] We will discuss the reasons why this is typically not the case in practice below.

perspective, after an innovation has taken place and after individuals know their place in the economy. Even though they may pretend to preach about idealized free markets, they clearly have not understood the full implications of how an idealized free market would work, i.e. that such a market would provide precisely the type of insurance that they are opposing.

In practice, even after they know that they are workers, the majority of workers replaced by technological progress do not have insurance contracts against being replaced. Of course there are good reasons for why such idealized risk markets are not present in the real world:

First, the limited lifespan of humans makes it difficult to write insurance contracts that stretch over multiple generations. Workers would have had to obtain the described insurance a long time ago, before AI was well-conceived and its implications were clear, when the associated insurance premium would have been commensurately low. Perhaps their far-sighted ancestors could have written state-contingent contracts on their behalf. Today, obtaining insurance against AI reducing wages would require workers to pay large amounts since the possibility is very real. In short, effective insurance would have had to take place behind a "veil of ignorance" about the likely advent of AI.
To put it another way, in this perspective, the first "insurable damage" to the individual occurs at the time that the probability of an innovation becomes non-negligible, for at that time, the insurance premium required for income smoothing becomes significant, and her welfare is lowered. The individual would have wanted to buy insurance against the risk that her insurance premium would go up. Thus, in a perfect market, insurance markets would have to go back at least to a date at which there was a negligible probability that the innovation occurs. This presents a problem: it may be that at the moment that the concept of AI is formulated precisely enough to be an insurable event (and therefore becomes an insurable event) it has a non-zero probability.

Second, even for more limited time periods, risk markets with respect to technological change are clearly not perfect. Among the main reasons are information problems:
Describing the State Space: This starts with the basic problem of how difficult it is to describe the future state space.[6] We cannot easily write a contract on something before it has been invented. Addressing this problem would require that an individual has to be insured against any technological event that leads to lower wages.
Furthermore, the curse of asymmetric information which inhibits insurance markets is as prevalent here as it is elsewhere:
Adverse Selection: Innovation leads to important adverse selection problems. Some people in the market are more informed than others—including than those who might provide insurance.

[6] Interestingly, this type of information problem is easy to deal with after innovation has occurred, because then we know what has been invented and in which state we are, but very difficult to capture in ex-ante contracts.

Although, there is no reason that workers would be more informed about the progress of AI than insurers, in an ideal market, the winners of innovation would provide insurance to the losers, and the winners (e.g. entrepreneurs) would almost certainly be better informed than the losers (e.g. workers).

Moral Hazard: Innovation may also be subject to moral hazard problems, i.e. the presence of insurance may affect the likelihood that the insured event occurs. Again, although workers are unlikely to affect the pace of innovation in AI, the actions of innovators may be, to some extent, affected. If they were to completely insure away all their returns from innovation, there would be scant incentive to exert effort.[7] Since, in a perfect insurance world, the winners would insure the losers, full insurance may be infeasible.

Insurance and redistribution

A natural counter-part to observation 1 is that in the absence of perfect insurance markets "behind the veil of ignorance," there generally is a need for redistribution. If workers have access to some insurance against the risk of AI but not perfect insurance, this doesn't remove the need for redistributions. Redistribution is generally needed unless the sale of insurance occurs "behind the veil of ignorance."

For example, obtaining AI insurance today would require workers to pay a large premium. Of course, conceptually, if one went back in time, before AI was well-conceived and its implications clear, one might argue that the premium would be low. But even that might not be so, since premia for large events, even with small probability, can be high. In any case, at the very moment of conception of AI—the first possible moment that one could conceivably have written a policy—AI would still have distributional consequences; workers would have to pay a premium to divest themselves of this risk, and thus they are worse off, and even more so, when compared to the innovators, the winners.

2.2. Perfect Markets Ex-Post and No Costs of Redistribution

Our next case pertains to a world that may be described as a 2nd-best world without the perfect insurance markets referred to earlier, but in which, *ex post,* all markets are functioning well *and* there can be costless redistributions. This case covers several critical results that, although

[7] Some might argue that this problem is equally hard to deal with before or after innovation has occurred. If we tax innovators ex-post, it destroys incentives just as much as if we fully insure away all returns from innovation. But this may not be so. If Bill Gates had been told, ex ante, that government would take away 50% of his returns over $10 billion, there is little reason to believe that it would have had any significant effect on innovation and investment. Ex post, taxing the winners in "winner takes all games" may have even less effect.

obvious at some level, often get lost in the debate about AI and technological progress more generally.

Observation 2) If redistribution is costless and appropriate redistributions are made, then technological progress is always desirable for all agents. In that case, there is political unanimity in supporting technological progress.

For convenience, and in conformity to conventional usage, we will refer the world with costless redistribution but otherwise perfectly functioning markets, as first best *ex-post*; though we remind the reader that the previous analysis suggested that in a true first best, workers would have insurance against the risk of AI, such that net income (after insurance) would be unaffected by the occurrence of the innovation. If the world is first-best *ex-post* in the sense thus defined, then the utility possibilities curve (or *Pareto* frontier) moves out. We provide an example in Figure 1, which depicts a utility possibilities frontier for two types of agents, workers and entrepreneurs. In the example, technological progress increases the maximum utility level of entrepreneurs for any given level of utility of workers.[8] Innovation has increased production possibilities, and with lump sum redistributions, an expansion in production possibilities automatically implies an expansion in utility possibilities, i.e. that everybody could be better off.

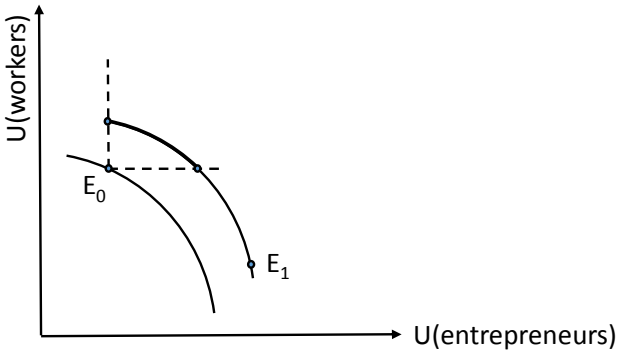

Figure 1: Pareto frontier before and after innovation with costless redistribution

The fact that they *could* be better off does not mean that they *will* be better off. That depends on institutional arrangements. In Figure 1, we denote the initial equilibrium by E_0 and the after-innovation equilibrium by E_1. We have deliberately not called it a competitive market equilibrium: markets don't exist in a vacuum (see e.g. Stiglitz *et al*, 2015). They have to be structured by rules and regulations, e.g. concerning intellectual property rights and anti-trust

[8] More generally, we could define a multi-dimensional utility possibilities frontier by adding any number of categories of individuals, or even naming the individuals.

policies, and there may be tax and other policies in place. We thus simply refer to E_0 and E_1 as the before and after innovation (institution-given) equilibrium. Note as drawn in the figure, workers are worse off. That would normally be the case with what Hicks referred to as labor-saving innovations, i.e. innovations which at a given wage, lead to less demand for labor. AI appears to be a labor saving innovation. In the simple formal models of worker-replacing innovations which we work out below, that is clearly the case.

This in turn has two important implications:

First, without adequate redistribution, it makes sense for workers to resist the innovation. Luddism – the movement named after the possibly fictional character Ned Ludd that opposed automation in the textile sector in late-18th and early-19th century England – is a rational response for workers who are worse off from automation and who are not sufficiently compensated.

Secondly, in a democracy in which workers are in a majority, it would make sense for enlightened innovators to support redistribution, to make sure that workers are at least not worse off. With redistribution, both innovators and workers *can* be better off. If appropriate redistribution is made so that everybody shares in the fruits of technological progress, there will again be political unanimity in supporting technological progress – progress will not be politically contentious.

There might be significant debate about how much compensation workers should receive, i.e. where in the "northeast corner of E_o" society should be. On the one hand, this debate concerns the distribution of the surplus generated by innovation. On the other hand, labor-saving innovation reduces wages, which generates a redistribution from workers to other factor owners like rentiers and capitalists, for which workers may seek compensation. This redistribution represents a pecuniary externality from the innovation, as we will discuss in further detail in Section 3.

In Figure 1, we have marked in bold that part of the post-innovation Pareto frontier which represents a Pareto improvement and lies to the northeast of E_o. A range of philosophical principles can be adduced for determining what is a "just" division of the fruits of innovation. Behavioral economics may provide insights into what kinds of divisions might be acceptable.[9]

[9] Consider a model in which workers and innovators have to agree on whether the innovation is acceptable. The innovator has the power to set the division of the gains (i.e. where along the curve Northeast of E_o the new equilibrium lies), but the workers has the power to accept or reject. This is the standard ultimatum game, for which there is a large body of literature suggesting that at least some of the fruits of innovation have to be shared with workers. If they perceive the allocation of benefits to be unfair, they would rather be worse off (e.g. at the

Of course, the innovation may not be labor saving, and the market equilibrium E_1 itself could be to the northeast of E_o. Although this case is easier, the distribution of the gains from innovation and any associated pecuniary externalities and rents may still be contentious, especially if they lead to large disparities in income. Distributive issues can also interact with production, as emphasized e.g. by the efficiency wage theories that we consider in greater depth in Section 5.

2.3. Perfect Markets but Costly Redistribution

There is another possibility—that as we try to redistribute, the new utility possibility curve may lie inside of the old utility possibilities frontier near the original equilibrium.

Observation 3) ***If the world is first-best (ex-post, after the innovation), but redistribution is limited or costly, then a Pareto improvement may not be possible, and some groups in society may oppose technological progress. With a sufficiently inequality-averse social welfare function, societal welfare may be reduced.***

This case is illustrated in Figure 2 below. The utility possibilities frontier is constrained by the costs imposed by redistribution. Even though it might appear that innovation could make everyone better off technologically, given the existing set of institutions of that economy, it actually can't – there may not be scope for avoiding utility losses for workers.

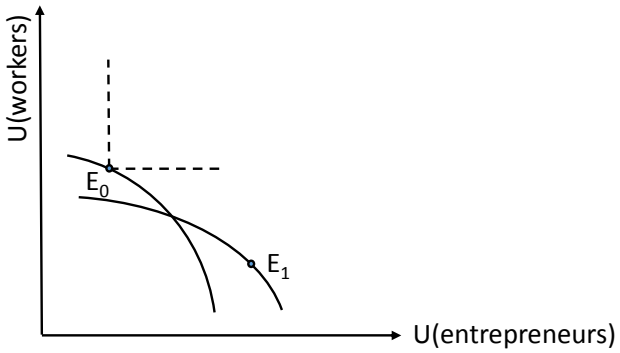

Figure 2: Potential Pareto frontier with costly redistribution

Some economists argue that the world looks like figure 2, and that if we try to transfer from innovators to workers, so much output is lost that workers are still worse off. If that is the case, then one cannot say that the innovation is a Pareto improvement. One hesitates to use the word "innovation." It is a change, perhaps a technological change, which has had the effect of

original point without the innovation) rather than at the point that just makes them indifferent to where they were before. See Fehr and Schmidt (2003).

making some people better off and others worse off. It is a distribution-inducing change and will be contentious.

A social welfare function that places no weight on inequality—which treats a dollar to rich innovators the same as a dollar to a poor worker—would, of course, conclude that the innovation is desirable. But with a more natural, inequality averse social welfare function, the so-called innovation is welfare decreasing.

The workers who lose out would rationally oppose the innovation. If workers are in a majority and innovators wish to maintain their position, it would behoove the innovators to think harder about how to engage in redistribution. This is, of course, a collective action problem for innovators – for individual innovators, the contribution to economy-wide inequality is typically limited, even if their collective behavior makes workers worse off. As a result, innovators often devote effort to actions which enhance their market power—lowering *real* incomes of workers still further—and to not paying taxes (both via clever tax avoidance using the existing legal framework, and via political lobbying to provide special exemptions from taxation for their industries). Disregarding, in our view unwisely, that their actions may stir up political opposition to innovation, some innovators go further and argue for weakening the progressivity of the tax system and a smaller state, so there are less public resources to provide for the well-being of the workers who are hurt by innovation.

According to a long-run version of "trickle-down" economics, repeated innovations will eventually increase the wealth of innovators so much that the benefits will trickle down to regular workers. In this view, a Pareto improvement is always possible in the long run, as in Figure 1, even if an entire generation of workers is hurt in the short- to medium run. This is a possibility and, in fact, the first industrial revolution may be an example. During the industrial revolution, workers eventually obtained enough human capital - which was publicly provided, as is in the interests of the innovators – so that the wages of almost all increased. However, once machines are smart enough, innovators may no longer have incentives to support the public financing of human capital accumulation, and it may well be that workers' standard of living decrease. In particular, in a political system dominated by money, the innovators, increasingly rich, may use their economic and political influence to resist redistribution. Furthermore, even if long-run trickle-down economics was correct, it may lead to tremendous suffering and social upheaval in the short run. It may also – understandably – not be very credible if innovators promise that once they are rich enough, they will support workers, but that they are not quite rich enough yet.

This leads to the important question: *How costly is redistribution in practice?* As we noted earlier, markets don't exist in a vacuum. They are structured by laws and regulations and how those laws and regulations are enforced. The outcome is the so-called "market" distribution of

income, which is then subject to taxes and transfer, leading to an after-tax distribution of income. But this conventional distinction may not be quite accurate: the rules of the game concerning redistribution affect the market income distribution. See Piketty et al. (2014) and Stiglitz (2017). The points that we have denoted E_o and E_1 describe the initial outcome and the outcome after-technological change, *assuming that laws, regulations, institutions, etc. remain unchanged.* But, of course, it is not reasonable to expect that they would remain unchanged with the advent of a change as significant as AI.

More particularly, each set of (feasible) laws, regulations, institutions, etc. defines a feasible utility possibilities frontier. We can think of the 2nd-best utility possibilities frontier as the outer envelope of all these frontiers. As the outer envelope, the 2nd-best utilities possibilities frontier provides more flexibility for redistribution than does that associated with one particular set of rules, regulations, and institutions. This reflects that any changes in laws, regulations, or institutions etc. will also have redistributive effects. Given this additional flexibility, the likelihood that a Pareto improvement as in Figure 1 can be achieved is greater. We provide further arguments for why this is likely to be the case in Section 3 (although we cannot entirely rule out a situation like that depicted in Figure 2 in which a Pareto improvement is not feasible).

2.4. Imperfect Markets

Let us also consider a fourth case, which does not necessarily reflect the specific situation with advances in artificial intelligence, but which is important to understand and keep in mind when we evaluate technological innovations.

Observation 4) If the economy is not first-best ex-post, then the utility possibilities frontier may move inwards in response to an expansion of production possibilities. Furthermore, this may even be true with costless redistribution.

When we speak about an economy that is not first-best, we mean an economy that deviates from the Arrow-Debreu benchmark, i.e. that exhibits market imperfections such as information problems, missing markets, price and wage rigidities which can result in aggregate demand problems, monopolies and monopsonies, and so forth. Typically, these mean that the market equilibrium is not Pareto efficient. The utilities possibilities frontier represents the maximum utility of workers, given that of entrepreneurs, taking the market failures as given.

This case is illustrated in Figure 3. The initial equilibrium is E_o, but the innovation, which would have led to greater efficiency in the absence of these market imperfections, makes workers worse off—and even with costless redistributions, there is no way that both workers and entrepreneurs can be better off.

An example, elaborated on by Deli Gatti *et al* (2012a,b), were the agricultural improvements at the end of the nineteenth century and beginning of the twentieth. The result was that agricultural prices plummeted, and so too did incomes on farms and in the rural sector. But mobility is costly—moving to the urban sector required capital, and many farmers saw their capital disappear as the value of their farms decreased. Those with loans often went bankrupt. Capital market imperfections (based on information asymmetries) meant that farmers couldn't borrow to move to the city to where the new jobs (hopefully) would be created. But as incomes in the rural sector plummeted, they couldn't buy the goods made by the manufacturing sector. Workers in both the rural and urban sector were worse off.[10] This provides at least one interpretation of the Great Depression—at least in the short run, these innovations proved Pareto inferior.

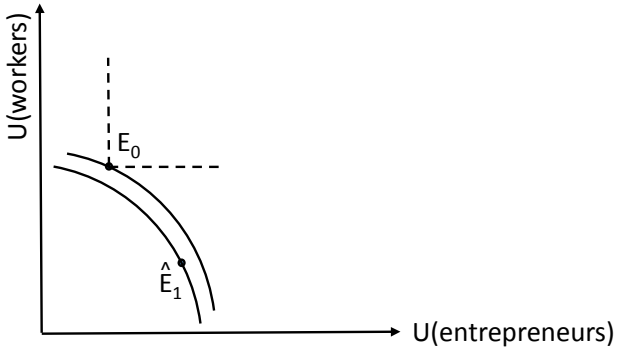

Figure 3: Potential Pareto frontier with market imperfections

So too, now standard results show that free trade may lead everyone to be worse off in the absence of good risk markets. (Newbery and Stiglitz, 1984). But that result can be interpreted as one involving technological progress. Assume that there were no way of transporting goods between the two countries. A technological advance allows goods to be transported freely. Then, under the quite plausible conditions postulated by Newbery-Stiglitz, welfare (of everyone!) in both countries could decrease.

The theory of the second best (Meade, 1955; Lipsey and Lancaster, 1956) reminds us that in the presence of market imperfections, improving the functioning of one market may deteriorate overall welfare. There are reasons to believe that certain innovations in financial markets, e.g.

[10] In the central Deli Gatti *et al* model, the agricultural sector has constant returns to scale and wages in the urban sector are rigid (e.g. because of efficiency wage considerations), so that the agricultural innovations are unambiguously welfare decreasing. In one variant of the model, where urban wages are flexible, wage decreases lead to still higher unemployment. Though it is possible that entrepreneurs gain more from the wage reductions than they lose from the loss of sales, social welfare is decreased with sufficiently inequality averse social welfare functions.

structured financial products and certain derivatives like credit default swaps, especially in the absence of appropriate regulations, contributed greatly to the Great Recession. (Financial Crisis Inquiry Commission (2011)).[11]

It is important to appreciate this result to understand how crucial our institutions and our market imperfections are in determining whether and how large a benefit society will derive from innovation.

Ascertaining whether the economy is in Case 1, 2, 3, or 4.

It is not always easy to ascertain which case best describes the economy. We observe only that the innovation has made some individuals better off, some worse off. (Here, we assume that it has made workers worse off, entrepreneurs better off.) The presumption is that redistributions are costly (so case 1 does not strictly apply) and markets are imperfect (so case 2 does not strictly apply.) But the costs of redistributions may be sufficiently low and the market imperfections sufficiently small that Figure 1 still applies: everyone could be made better off. Alternatively, redistributions may be so costly that Figure 2 applies. Or market failures are sufficiently large and redistributions sufficiently costly that Figure 3 applies.

We emphasize that which situation we are in depends not just on the possibilities of *ex post* redistribution, but on the institutional flexibility, which determines the ex ante distribution.

As we noted, the 2nd-best utilities possibilities frontier is the outer envelope of all conceivable constrained utilities possibilities frontiers, which reflect all the conceivable institutional regimes in an economy and all the market imperfections that the economy may potentially suffer from. By *institutional regimes* we mean all explicit tax and redistribution systems (from negative income tax systems to universal basic income to the regressive tax system currently in place), intellectual property regimes, job programs, education programs, but even social norms such as those related to charitable contributions. *Market imperfections* include all the market arrangements that differ from the Arrow-Debreu "optimal" benchmark, the conditions which ensure the Pareto efficiency of the market. As we noted earlier, the term embraces imperfections in information, competition, and risk and capital markets (including "missing" markets) but also rigidities in factor reallocation or in prices that determine how easily factors and products reallocate and which may be particularly important in the context of technological progress.

Changing any of these institutions or market imperfections has an effect on workers' welfare. In general, it may be desirable to use a package of changes to all these institutions to ensure

[11] At a theoretical level, Simsek (2013) and Guzman and Stiglitz (2016a, 2016b) have shown that opening up new markets—through financial innovation-- can lead to greater volatility in consumption.

Pareto improvements after technological change has occurred. For instance, in Section 4.3, we show that a combination of a change in the intellectual property regime *and* a change in capital taxation can ensure that an innovation is a Pareto improvement.

Finally, we also note that the possibility of achieving a Pareto improvement depends on how broadly we define the classes of individuals which are affected by an innovation. Our earlier example distinguished society, for simplicity, into two categories, workers and entrepreneurs. More generally, different categories of workers, e.g. skilled and unskilled workers, or workers in different sectors or tasks, are differentially affected by innovation. By the same token, different categories of entrepreneurs or innovators are differentially affected by innovation – for example, a given entrepreneur will generally be worse off if she is out-competed by another's innovation. In the limit, if we consider the welfare of every single agent in the economy, clear Pareto improvements in a strict sense will be very difficult to find. As a result, our scope of analysis has to be targeted at the level that is relevant for the question at hand.

From both a political and a macroeconomic perspective, it is desirable that our welfare analysis focuses on groups that are sufficiently broad so that they matter for the political or economic equilibrium. It may also be useful to focus on groups that can be targeted with specific policy measures. Having said that, there is also a useful role for social safety nets that insure single individuals that lose out – for example, an innovator who goes broke because he was outpaced by a competitor.

2.5. Endogenous Technological Progress

A last and fifth point to emphasize is that there is no 1st welfare theorem for endogenous innovation. Generally speaking, the private returns to innovation in an economy differ from the social returns.[12]

Observation 5) The privately optimal choice of innovation may move the utility possibilities frontier inwards, even if redistribution is costless and the economy is ex-post first-best.

This implies that there may be benefits from intervening in the innovation process to generate Pareto improvements, for example, by making it less labor saving (see e.g. Stiglitz, 2014b). Again, this does not specifically refer to advances in artificial intelligence – it will probably not

[12] It is hard to know who first had this insight. Certainly, Thomas Jefferson, America's third president, recognized it when he said that knowledge is like a candle: when it lights another, the light of the first candle is not diminished. In the economics literature, it was clearly articulated by Arrow (1962) and Stiglitz (1987a). For a more recent statement of why social and private returns to innovation differ, see Stiglitz and Greenwald (2015). These results hold regardless of the intellectual property regime. Poorly designed intellectual property regimes can (and do) impair innovation. For a simple theoretical model, see Stiglitz (2014a); for empirical evidence, see Williams (2013).

apply to most examples of innovation in AI – but it is easy to bring examples where privately optimal innovation may generate Pareto deteriorations, for example in the context of high-frequency trading in financial markets (see e.g. Stiglitz (2014c)).

2.6. Relationship between Technological Progress and Globalization

Many of the effects of technological change in general and AI in particular are similar to those of globalization. Indeed, globalization can be viewed as a change in technology, that of trading with the rest of the world. In particular, trade of advanced countries with developing countries is "labor saving" (in the sense of Hicks): the demand for unskilled workers, or workers in general, decreases, at any given wage, implying that while the production possibilities curve moves out, and the utilities possibilities curve may move out, the new equilibrium entails workers being worse off, as in Figures 1 and 2. (In the absence of good risk markets, as we noted, everyone can be worse off, as in Figure 3). Thus, the issue of whether globalization is welfare enhancing comes back to the question addressed in this paper: is it possible to ensure, either through redistributive taxes or changes in institutions/rules, that workers are not made worse off. Again, there is a presumption that the gains to capital (or enterprises) could be taxed, to provide the requisite redistributions.[13]

As we discuss in greater detail below, one of the side effects of innovation and IPR is the creation of market power. Similarly, one of the consequences of globalization is to weaken the market power of workers. This is important because there is ample evidence that labor markets are far from perfectly competitive. The requisite compensation and/or offsetting changes in institutional rules to ensure that globalization represents a Pareto improvement may thus have to be all the greater.

3. Technological Progress and Channels of Inequality

There are two main channels through which technological progress may affect the distribution of resources and thus inequality: first, through the surplus earned by innovators and secondly, through effects on other agents in the economy.

[13] Although a given country that opens up to trade is always made better off in a first-best world, ensuring a global Pareto improvement after a country reduces its trade barriers may be even more difficult than after technological progress has occurred, since changes in trade barriers affect international terms of trade and lead to redistributions across all other countries that can only be undone via cross-country transfers. See e.g. Korinek (2016). Furthermore, within each country, gains from trade inherently require changes in relative prices, which means that large redistributions are even more likely than in the case of technological progress.

3.1. Surplus Earned by Innovators

Technology is an information good, which implies that it is *non-rival*, but it may be *excludable*. Non-rival means that information can be used without being used up – in principle, many economic actors could use the same technology at the same time. If information about an innovation is widely shared, it can be used by all of society and provide welfare benefits to all. The excludable nature of information means, however, that others can be prevented from either obtaining or using a technology, for example by withholding it from the public (e.g. as a business secret) or by using social institutions such as intellectual property rights (e.g. copyrights or patents). This excludability may provide innovators with market power that enables them to charge a positive price for the innovation and earn a surplus.

Society faces a difficult trade-off in determining how to engineer the optimal level of innovation. In a first-best world, there are no agency problems in the process of innovation, and an optimal solution would be for the public to fund innovations and make them freely available to all (see e.g. Arrow, 1962). In fact, this model of financing innovation is common for basic research and has given rise to some very significant innovations in history, including the invention of the Internet. A closely related solution is the production of innovations for non-pecuniary rewards, such as e.g. the prevalence of open source technology, which is widespread in the context of software and even artificial intelligence.[14]

However, in many circumstances, private agents are superior in producing innovation, and when they fund innovation, they expect to earn a return. The surplus earned by innovators then plays an important economic role because it rewards innovators for what they accomplish – it represents the economic return to innovation activity. However, there is often market power associated with innovations, especially when there is a system of IPR in place, and this generally leads to inefficiencies compared to the first-best allocation in which innovations are distributed as public goods.[15]

We distinguish the following two cases, which determine whether innovators earn rents, i.e. payoffs in excess of the cost of their innovative activity:

First, if entry into innovation activity is restricted, then the surplus or net income earned by innovators is generally greater than the costs of innovation activity. A natural example of such

[14] This approach relies on individuals or companies that are willing to innovate in exchange for non-pecuniary rewards such as prestige or, alternatively, on a calculated decision that providing free technology will steer potential customers or employees towards an innovator's platform, as seems to be the case in the field of AI.
[15] For discussions of the merits of alternative ways of funding and incentivizing innovation, see Dosi and Stiglitz (2014), Baker *et al* (2017), Stiglitz (2008) and Korinek and Ng (2017).

restrictions is when only a small number of people are endowed with special skills that enable them to innovate. These innovators then earn rents based on their exclusive abilities.

Restrictions to innovative activity may also arise from market structure: in markets with Bertrand competition, the first entrant who develops a costly innovation may enjoy a monopoly position because any potential competitor knows that if she enters, the incumbent will cut prices to marginal cost so that she cannot recoup the investment into an innovation.[16]

Secondly, if innovative activity is contestable, i.e. if there is a sufficiently large set of potential innovators with equal skills, then the expected rents to innovative activity are competed down to zero, i.e. the marginal entrant into innovative activity is indifferent between innovating or not.[17] However, given that the payoffs to innovation are highly stochastic, there will be winners and losers ex-post. In the context of new technologies, the distribution of payoffs seems to be increasingly skewed, with a small number of entrepreneurs earning gigantic payoffs and the vast majority earning little in return for their efforts. This gives rise to significant inequality even among innovators.

In either case, the returns earned by an innovator may not correspond closely to the social returns to the innovation; in particular, some of the returns may reflect the capture of profits that would otherwise have gone to other entrepreneurs.

Policies to Share the Surplus of Innovators

There is a growing consensus that one of the sources of the growth of inequality is the growth of rents, including the rents that innovators earn in excess of the cost of innovation (see e.g. Korinek and Ng, 2017). Taxing and redistributing such rents has an important role in ensuring that AI and other advances in technology are Pareto improving. Also, anti-trust policies may lower such rents, ensuring that the benefits of innovations are more widely shared, as more competition lowers consumer prices from which all benefit. From the perspective of low skilled workers who losing from innovation, targeted expenditure programs financed by high rent taxes may be of greater benefit than the lowering of prices, the benefit of which will go disproportionately to those who otherwise have spending power.

---

[16] See Stiglitz (1987b) and Dasgupta and Stiglitz (1988). When the number of firms is limited and there is Cournot competition, there will also be rents associated with innovation. For more general theoretical discussions of industrial structure and innovation, see Dasgupta and Stiglitz (1980ab) and Stiglitz and Greenwald (2015, ch. 5).

[17] Given the difficulty of predicting the success of innovative activity or of even assigning success probabilities, it is questionable how efficiently this mechanism works in practice. For example, there may be excessive entry because of over-optimism by some potential entrepreneurs, or there may be insufficient entry because of imperfect insurance markets for entrepreneurs.

If some are better at innovating than others (and know it), then these individuals will enjoy infra-marginal innovation rents (on average.)

Moreover, changes in intellectual property rights (IPRs) affect who receives the benefits of innovation—and thus the "incidence" of innovation, since IPRs are instrumental in providing extended market power to innovators.[18]

Additionally, public research – with government or the public at large appropriating the returns, rather than allowing private firms to do so as now – together with stronger competition policies, might reduce the scope for monopolies capturing large fractions of the returns to innovation, and thus enhance the likelihood that AI will be Pareto improving.

Workers may also note that the innovations that ultimately led to AI – including those created by private entrepreneurs – build on significant public support. Society as a whole, but not necessarily this generation of innovators, paid for this knowledge, and should therefore share in the surplus generated by the innovation. One proposal to ensure that workers share in the benefits of innovation—and are less likely to lose from it—is to give workers shares in enterprises to ensure that their welfare goes up in tandem with that of shareholders/innovators as a whole.

3.2. Effects on Others

Innovation also leads to large redistributions among others in the economy who are not directly involved in the process of innovation, for example workers who experience a sudden increase or decline in the demand for their labor. These redistributions can thus be viewed as externalities from innovation, and they are one of the main reasons why innovation raises concerns about inequality. We distinguish two categories of such externalities, pecuniary and non-pecuniary externalities. We discuss both in detail in the following:

Pecuniary Externalities: Price and Wages Changes

Among the most prominent implications of technological change is that it affects the prices of factors of production (including wages) and of produced goods. Hicks (1932) already observed that innovations generally change the demand for factors and will, in equilibrium, lead to factor price changes, especially changes in wages.

If – as many technologists predict – artificial intelligence directly replaces human labor, the demand for human labor will go down, and so will wages. More generally, innovations typically reduce demand for specific types of labor with specific human capital. For example, self-driving cars will likely depress the wages of drivers, or radiology-reading AI may lower the wages of traditional radiologists. Conversely, AI has certainly led to an increase in demand for computer

[18] Especially when there is Bertrand competition, the benefits of innovation may be quickly shared with consumers upon the termination of patents.

scientists and has greatly increased their wages, in particular in sub-fields that are directly related to AI. Since AI is a general purpose technology, there are reasons to believe that advances in AI will reverberate throughout many different sectors and lead to significant changes in wages throughout the economy in coming decades. Similar arguments can be made about the demand for and the value of different types of specific capital, as well as the demand for and prices of particular products.

Even though there are frequently losers, technological progress by definition shifts out the production possibilities frontier. This implies that the total dollar gain of the winners of progress exceeds the dollar loss of the losers lose[19]. In section 4 below, we will use this property of technological progress to argue that, under relatively broad conditions, this should enable the redistribution that is necessary to ensure that innovation leads to a Pareto improvement: the gains that arise to some factor owners as a result of technological progress are excess returns that are like unearned rents and could be taxed away without introducing distortions into the economy.

The pecuniary externalities from innovation give rise to the need for potentially costly redistributions, and they make the economy inefficient relative to the idealized insurance setting that we analyzed in section 2: if "behind the veil of ignorance" risk markets are absent, then redistribution is a substitute for missing insurance markets. Redistribution would insure workers and the owners of human and physical capital against such pecuniary externalities.

Policies to Counter Wage Declines

There are a range of further policies to counter the wage declines that are experienced by workers who are displaced by machines, even for low skilled jobs. These include wage subsidies and earned income tax credits. If bargaining power in labor markets is biased towards employers, an increased minimum wage can also help ensure that no one who works full time is in poverty. Furthermore, ensuring high aggregate demand—and thus a low unemployment rate—also increases the bargaining power of workers and leads to higher wages.

Other policies aimed at increasing the demand for especially low skilled labor include higher wages in the public sector as well as an increase in public investments and other public expenditures; all of these policies help to drive up wages in the economy more generally.

Policies that could be used to finance such measures include carbon taxes, which would encourage resource saving innovation, at the expense of labor saving innovation. It would thus

[19] If lump sum transfers were feasible, the winners could compensate the losers. This does not mean that social welfare is higher in the absence of such compensation.

simultaneously address two of most serious global problems, global climate change and inequality.[20]

Furthermore, the elimination of tax deduction for interest and the imposition of a tax on capital would increase the cost of capital and induce more capital augmenting innovation rather than labor saving innovation.

Non-pecuniary Externalities

Innovation may also generate non-pecuniary externalities on agents other than the innovator. Classic examples for this are technological externalities – for example, if an innovation produces public goods, or generates or alleviates pollution. In markets that deviate from the Arrow-Debreu benchmark, a variety of non-pecuniary effects may arise: for example, innovation may affect quantities demanded, or the probability of buying or selling a good or factor, including the probability of being unemployed.

Some effects are such that they can be interpreted either as pecuniary or non-pecuniary externalities. For example, product innovations can be interpreted as a price changes – the price of the newly invented good changes from infinity to some positive value – or as a change in the price of the consumption services provided by the good. Alternatively, they can also be interpreted in a non-pecuniary manner by viewing a product (such as a smartphone) as providing a bundle of services to consumers which can only be bought in fixed proportion (e.g. since we cannot separately purchase different functions of the smartphone). In that view, an innovation represents a change in the structure of incomplete markets because it changes the bundle of consumption service available from a product. Similarly, changes in job quality can be interpreted by viewing each job as a vector of transactions that are only available in pre-determined bundles, and the innovation changes the elements in the bundle that are available. It is well-known, that changes in the degree of market incompleteness for such bundles give rise to externalities (a specific application of Greenwald and Stiglitz, 1986).

4. Worker-Replacing Progress and Redistribution

This section considers a stark form of technological progress that we term worker-replacing technological progress. We develop two simple models to analyze the two channels generating inequality that we discussed in the previous section. In Sections 4.1 and 4.2, we consider the pecuniary externalities (redistributions) generated by worker-replacing progress, both from a static and a dynamic perspective. In Section 4.3., we focus on the distribution of the surplus

[20] As we noted above, there is no first fundamental welfare theorem for innovation, and indeed, there is a presumption that the market is biased towards labor saving innovation relative to innovations directed towards "saving the planet." See Stiglitz 2014b.

accruing to innovators in a model in which the surplus is determined by the level of patent protection. Furthermore, in Section 4.4 we discuss the implications of endogenous factor bias in technological progress.

4.1. Static Pecuniary Externalities of Worker-Replacing Progress

For sections 3.1 and 3.2, we consider the simple model of worker-replacing technological change of Korinek and Stiglitz (2017). We assume a production technology that combines capital and labor in a constant-returns-to-scale (CRS) function, but where labor consists of the sum of human and machine labor. Assuming that human and machine labor enter the production function additively means that they are perfect substitutes for each other. The details of the baseline model are presented in Box 1 below.

We analyze three questions: What does worker-replacing technological change do to wages in the short-run and in the long-run? And what can policy do about it?

First, we look exclusively at the short-run before any of the other factors have adjusted:

Observation 1) ***Machine Labor and Factor Earnings (in the short-run):*** *adding a marginal unit of machine labor reduces human wages but increases returns of complementary factors in a zero-sum manner.*

Intuitively, what happens if we add one unit of machine labor is that first, that unit will earn its marginal return, but secondly, there is also a redistribution from labor to capital, which now becomes relatively scarcer. The gains of capital are exactly the losses of the existing stock of labor.

The redistribution generated by technological progress can be thought of as a pecuniary externality, as we emphasized earlier. The income losses of wage earners and the income gains of other factors owners are inefficient compared to the first-best benchmark considered in Section 2.1. In the given example, the owners of capital have obtained windfall gains but have not done anything to earn these higher return. A compensatory transfer from capital owners to workers simply undoes these windfall gains and leaves them equally well off as they were before.

More generally, adding machine labor creates a redistribution away from human labor toward complementary factors. More generally, the result holds no matter what the complementary factor, for instance whether it is capital or land or unskilled versus skilled labor or entrepreneurial rents. Policy can undo these redistributions by taxing windfall gains while leaving the price system to work at the margin. The result also holds for decreasing-returns-to-scale production functions if we interpret the profits earned by the owner of the technology as

Box 1: Machine Labor and Factor Earnings

Assume a constant-returns-to-scale production function that produces output Y by combining capital K with labor, consisting of the sum of human labor H and machine labor M.

$Y = F(K, H + M)$

In this formulation, human labor and machine labor are perfect substitutes so machine technology is clearly what we call worker-replacing.

In the competitive equilibrium, the wage is determined by the marginal product of labor,

$w = F_L$

Proposition 1: Machine Labor and Factor Earnings: adding a marginal unit of machine labor reduces human wages but increases the returns to capital in a zero-sum manner, in addition to increasing output by the marginal product of labor, which is equal to the wage.

Proof: Using Euler's Theorem, we rewrite the production function:

$$(H + M)F_L(\cdot) + KF_K(\cdot) = F(K, H + M)$$

We can now ascertain the effect of an additional unit of M:

$$F_L + (H + M)F_{LL} + KF_{KL} = F_L$$

or simplified:

$$\underbrace{(H + M)F_{LL}}_{decline\ in\ wage\ bill} + \underbrace{KF_{KL}}_{increase\ in\ return\ to\ K} = 0$$

Source: Korinek and Stiglitz (2017)

compensation for the implicit factor "entrepreneurship," which takes part in the zero-sum redistribution.

Let us also emphasize that taxes on previously accumulated factors that suddenly earn an unexpected excess return are non-distortionary. This means that at least in principle, there is a role for implementing costless redistribution and generating a Pareto-improvement. (In practice, there are some natural caveats to this result. For example, it relies on the assumption that we can distinguish between previously installed capital that earns windfall gains and new capital that would be distorted if it were taxed.)

4.2. Dynamic Implications of Worker-Replacing Progress

In the longer run, worker-replacing technological change will lead to significant economic change. It implies that the biggest constraint on output – the scarcity of labor – is eventually, and suddenly, lifted. As a result, greater amounts of complementary factors, here capital, are accumulated.

Observation 2) ***Machine Labor and Abundance of Labor:*** *If not only capital but also labor is reproducible at sufficiently low cost, then the economy will grow exponentially in AK-fashion, driven purely by factor accumulation, even in the absence of further technological change.*

In Korinek and Stiglitz (2017), we describe the dynamics of this transition, as machines made by machines get increasingly efficient or equivalently, as the cost of producing machines decreases. We identify a singularity point at which it becomes cost-effective for machines to start to fully replace human labor.[21] In the simplest case, when complementary factors such as capital adjust without friction, the human wage may actually be unchanged because capital K grows in proportion to effective labor (H + M) so that the marginal productivity of labor and the wage remain unchanged. In other words, investment is allocated between conventional machines and human-replacing robots in such a way that the return is equal to the intertemporal marginal rate of substitution. Under the assumption that workers only care about their absolute income, not their relative income, this outcome would not be too bad for workers: in absolute terms, even though the human labor share would go to zero as an increasing fraction of the labor in the economy is performed by machines, workers are no worse off as a result of AI.

When factors are slow to adjust, the pattern of transition can be complex, with demand for human labor typically going down temporarily.[22] In general, the pattern of adjustment depends on how fast the capital stock versus the stock of labor adjust. (For example, if the capital stock rises in anticipation of an increased supply of machine labor in the future that has not yet materialized, then human wages may even go up at intermediate stages.[23])

[21] This singularity captures the important economic aspects of what technologists such as Vernor Vinge (1993) or Ray Kurzweil (2005) call the technological singularity. A similar point is also made in Aghion et al. (2017).
[22] Berg et al. (2017) shows that it may actually take decades for the economy's complementary capital stock to adjust after major revolutions in labor-saving technology.

[23] This assumes that capital is "putty-putty," i.e. that capital investments made before AI arrives are equally productive after AI, as would be the case if humans and robots were in fact identical.

However, the following observation describes that workers are actually worse off as a result of machine labor if there are non-reproducible complementary factors that are in scarce supply, such as land or other natural resources:

Observation 3) ***Machine Labor and Return of Scarcity:*** *if there are non-reproducible complementary factors, they eventually limit growth; human real wages fall, and the owners of non-reproducible factors absorb all the rents.*

Intuitively, as the supply of effective labor proliferates due to the introduction of machine labor, agents in the economy will compete for scarce non-reproducible resources like land, driving up their price.

A similar argument holds for non-reproducible consumption goods: even if all factors in the production process are reproducible so that productive output in the economy exhibits AK-style growth and workers' product wages remain unchanged, competition for fixed resources that are part of their consumption basket, such as land used for housing, may lead workers to eventually be worse off. This may be particularly important in urban settings where, say, economic activity occurs at the center. Rich rentiers may occupy the more desirable locations near the center, with workers having to obtain less expensive housing at the periphery, spending more time commuting. The advent of AI will thus lower their utility.

However, just as in the earlier case, at the margin, the redistribution from workers to non-reproducible factor owners is zero sum. Since taxes on non-reproducible factors are by definition non-distortionary, there is scope for non-distortionary redistribution.

Observation 4) ***Non-Reproducible Factors and Pareto Improvements:*** *So long as non-distortionary taxes on factor rents are feasible, labor-replacing innovation can be a Pareto improvement.*

4.3. Redistributing the Innovators' Surplus via Changes in Institutions

If outright redistribution is infeasible, there may be other institutional changes which result in market distributions which are more favorable to workers. For example, intervention to steer technological progress may act as a 2nd-best device.

In this section, we provide an example in which a change in intellectual property rights—a shortening of the term of patent protection—effectively redistributes some of the innovators' surplus to workers (consumers) to mitigate the pecuniary externalities on wages that they experience, with the ultimate goal that the benefits of the innovation are more widely shared. If an innovation results in a lower cost of production, then the innovator enjoys the benefits of

Box 2: Intellectual Property Regime and Redistribution

Consider an economy with a unit mass of workers H = 1, in which the capital stock supplied each period K(τ) is a function solely of a distortionary capital tax τ, the proceeds of which are distributed to workers, and the effective stock of machine labor M(z) is an increasing function of patent life z.

A worker's total income I consists of her wage plus the revenue of the capital tax,

$$I = w + \tau K(\tau)$$

For any level of M(z), we define τ(M) as the value of the capital tax that keeps workers just as well off as they were before the introduction of machine labor.

Proposition 1 As long as elasticity of capital supply is not too large, we can always increase z from z = 0 and compensate workers by raising the capital tax τ.

Steady state dynamics

Consider a intertemporal setting in which the growth rate g = g(z, τ) is a function of the length of the patent z and the tax rate τ on innovators. Assume that the share of output that is invested is a function of the growth rate (i(g)) and that the fraction of output not spent on investment that is appropriated by the innovator is b(z, τ). In steady state, the present discounted value of the income of workers can be approximated as

$$PDV = (1 - i(g))\,[1 - (1 - \tau)b(z, \tau)]/(r - g)$$

where r is the discount rate. If we choose {z, τ} to maximize the PDV, in general, the optimum will not be a corner solution in which innovation hurts workers.

Proposition 2 In general, the optimal {z*, τ*} entails g > 0.

It is easy to write down sufficient conditions under which proposition 2 holds: setting τ* equal to zero, all that we require is that $|g_z|$ is not too large relative to $|b_z|$.

the innovation in the form of higher profits during the life of the patent; but after the expiration of the patent, society enjoys the benefits in terms of lower prices. The trade-off is that shortening the life of the patent may reduce the pace of innovation. But in the spirit of the theory of the second-best, there is generally an "optimal" patent life, in which there is still some innovation, but in which the well-being of workers is protected.

With network externalities, even after the end of the patent, the innovator may be able to maintain a dominant position after the end of the patent, and may continue to earn the surplus from her innovation. With taxes on monopoly profits, it should be possible to ensure that the innovations are Pareto improving and that even human worker-replacing technical change can improves the well-being of workers.

4.4. Factor biased technological change

So far, we have simply assumed that technological change – the introduction of AI – is worker-replacing. But advances in technology also make machines more productive, lowering the return to traditional capital[24]. It is thus useful to think of the world as having three groups: capitalists, workers, and innovators. Intellectual property rights (and anti-trust laws) determine the returns to innovators; but the nature of technological change in a competitive market determines the divisions between workers and capitalists.

A long-standing literature, going back to Kennedy (1964), Von-Weizacker (1966), and Samuelson (1965), describes the endogenous determination of the factor bias of technological progress.[25] The central result is that as the share of labor becomes smaller, the bias shifts towards capital-augmenting technological progress. If the world works as these models suggest, this should limit the decline in share of labor (at least in a stable equilibrium) and in inequality.[26] As the share of labor decreases, the incentive to produce worker-replacing innovation such as AI decreases. But the relevant discounted future wage share near the point of singularity may be sufficiently great that there is nonetheless an incentive to pass the point of singularity.

Let us assume that land becomes the binding constraint once human labor is fully replaceable by machine labor. In that case, provided the elasticity of substitution between land the other production factors – capital cum labor – is less than unity, the share of land increases over time, generating the result (analogous to that where labor is the binding constraint in the standard literature) that in the long run, all technological progress is land-augmenting. If the production function is constant returns to scale in land, labor (including machine labor) and traditional capital, then the long run rate of growth is determined by the pace of land-augmenting technological change.

[24] As we noted earlier, IA (intelligence assisting) innovation may increase the productivity of humans, and thus increase the demand for humans.

[25] Important contributions were also made by Drandakis and Phelps (1965). More recently, there has been some revival of the literature, with work of Acemoglu (2002), Stiglitz (2006, 2014b) and Acemoglu and Restrepo (2016), among others.

[26] One can describe dynamics with standard wage-setting mechanisms. The system is stable so long as the elasticity of substitution between factors is less than unity (Acemoglu, 1998; Stiglitz, 2006, 2014b).

Role of the Service Sector

Currently, progress in AI focuses on certain sectors of the economy, like manufacturing. Partly because of the resulting lower cost of manufacturing, partly because of the shape of preferences, the economy is evolving towards a service sector economy. (If there is differential productivity across sectors, and the elasticity of demand for the innovation sector is not too high, then production factors will move out of that sector into other sectors. This is even more so if preferences are non-homothetic, e.g. demand for food and many manufactured goods having an income elasticity less than unity.) Among the key service sectors are education, health, the military, and other public services. The value of those services is in large part socially determined, i.e. by public policies not just a market process. If we value those services highly—pay good wages, provide good working conditions, and create a sufficient number of jobs—this will limit increases in the inequality of market income. Governments typically play an important role in these sectors and their employment policies will thus play an important role in the AI transition. Many of these service sector jobs have limited skill requirements. However, higher public sector wages services will – through standard equilibrium effects – also raise wages in the private sector, will improve the bargaining position of workers, and will result in such jobs having higher "respect." All of this will, of course, require tax revenues. If the elasticity of entrepreneurial services is low, we can impose high taxes to finance these jobs.

5. Technological Unemployment

Unemployment is one of the most problematic societal implications of technological progress– new technology often implies that old jobs are destroyed and workers need to find new jobs. Economists, of course, understand the "lump-of-labor fallacy" – the false notion that there is a fixed number of jobs, and that automating a given job means that there will forever be fewer jobs left in the economy. In a well-functioning economy, we generally expect that technological progress creates additional income, which in turn can support more jobs.

However, there are two sound economic reasons for why technological unemployment may arise: first, because wages do not adjust for some structural reason, as described e.g. by efficiency wage theories, and second as a transition phenomenon. The two phenomena may also interact in important ways, e.g. when efficiency wage considerations slow down the transition to a new equilibrium. We discuss the two categories in turn in the following subsections.

The unemployment implications are especially problematic when technological progress is labor-saving, which – by definition – requires that either wages have to fall or that other complementary factors like capital have to adjust enough for labor market equilibrium to be restored.

5.1. Efficiency Wage Theory and Non-Adjustment of Wages

The first category of technological unemployment arises when wages do not adjust for structural reasons. Efficiency wage theory emphasizes that productivity depends on wages and so employers may have reasons to pay wages above the market clearing level. The original efficiency wage paper (Stiglitz, 1969) noted one of the reasons for this: that income disparities can weaken worker morale. Akerlof and Yellen (1990) have formalized this into the "fair wage hypothesis."

If fairness considerations are significant enough, and workers think that a decrease in their wages is "unfair" (for example because the income of entrepreneurs increases so entrepreneurs could easily "afford" pay increases), it means that the scope of labor-saving progress which shifts the utility possibilities curve out *without redistributions* is very limited. The new utility possibilities curve may lie outside the old one to the "north" of E_o, i.e. there is scope for a Pareto improvement in principle; but it may lie inside of the old utility possibilities curve near E_1, i.e. the utility possibilities of workers decrease for a given level of utility of entrepreneurs because workers reduce their effort so much that the effective labor supply declines—any gains from technology are more than offset by increased shirking. Shapiro and Stiglitz (1984) emphasize that paying a wage above the market-clearing level reduces shirking, leading to unemployment.

An even more daunting example of efficiency wages may arise if automation continues and the marginal product of labor for low-skill workers falls below their cost of living (even if they exert their best effort). Unless basic social services are provided to such workers, a nutritional efficiency wage model applies in that case, similar to what Stiglitz (1976) described for developing countries: employers could not pay a market clearing wage because they know that this would be insufficient for their employees to provide for themselves and remain productive.[27] We will follow up on this theme in the final section of our paper.

In traditional efficiency wage models, the unemployment effects of efficiency wages are permanent. For example, if technological change leads to greater inequality (or better information about the existing level of inequality), morale effects and the resulting efficiency wage responses imply that the equilibrium level of unemployment rises.

However, efficiency wage arguments may also contribute to slowing down the transition to a new equilibrium after an innovation, as we will explore in the next subsection.

[27] Even worse outcomes could emerge in the presence of imperfect capital markets, if expenditures on health and nutrition at one date affect productivity at later dates.

Minimum Wages and Non-Adjustment of Wages

An alternative reason why wages may not adjust to the market-clearing level are minimum wage laws. Basic economics implies that there will be unemployment if wages are set to an excessive level. Although this is a theoretical possibility, recent experience in the US has repeatedly shown that modest increases in minimum wages from current levels have hardly any employment effects but raise the income of minimum wage workers, which may have positive aggregate demand effects since low-income workers have a high marginal propensity to consume (see e.g. Schmitt, 2013). From an economic theory perspective, these observations are possible because wages are not determined in a purely Walrasian manner – there is a significant amount of bargaining involved when prospective employers and employees match – and increases in minimum wages substitute for the lacking bargaining power of workers (see e.g. Manning, 2011).

5.2. Technological Unemployment as a Transition Phenomenon

The second category of technological unemployment is as a transition phenomenon, i.e. when technological change makes workers redundant at a faster pace than they can find new jobs or that new jobs are created. This phenomenon was already observed by Keynes (1930). It is well understood that there is always a certain "natural" or "equilibrium" level of unemployment as a result of churning in the labor market. In benchmark models of search and matching to characterize this equilibrium level of unemployment (see Mortensen and Pissarides, 1994 and 1998), employment relationships are separated at random, and workers and employers need to search for new matches to replace them. The random shocks in this framework can be viewed as capturing, in reduced form, phenomena such as lifecycle transitions but also technological progress in individual firms. In this view, an increase in the pace of technological progress corresponds to a higher job separation rate and results in a higher equilibrium level of unemployment.

The transition may be especially prolonged if technology implies that the old skills of workers become obsolete and they need to acquire new skills and/or find out what new jobs match their skills (see e.g. Restrepo, 2017).

Even if in the long run, workers adjusted to AI, the transition may be difficult. AI will impact some sectors more than others, and there will be significant job dislocation. As a general lesson, markets on their own are not good at structural transformation. Often, the pace of job destruction is greater than the pace of job creation, especially as a result of imperfections in capital markets, inhibiting the ability of entrepreneurs to exploit quickly new opportunities as they are opened up, which has arguably been the case for globalization in many developing countries.

The Great Depression as an Example of Transitional Unemployment

The Great Depression can be viewed as being caused by rapid pace of innovation in agriculture (see Delli Gatti et al., 2012a). Fewer workers were needed to produce the food that individuals demanded, resulting in marked decline in agriculture prices and income, leading to a decline in demand for urban products. In the late 1920's, these effects became so large that long standing migration patterns were reversed.

What *might* have been a Pareto improvement turned out to be an immiserizing technological change, as both those in the urban and rural sector suffered.

The general result is that noted earlier: with mobility frictions and rigidities (themselves partly caused be capital market imperfections, as workers in the rural sector couldn't obtain funds to obtain the human capital required in the urban sector and to relocate) technological change can be welfare decreasing. The economy can be caught, for an extensive period of time, in a low level equilibrium trap, with high unemployment and low output.

In the case of the Great Depression, government intervention (as a by-product of World War II) eventually enabled a successful structural transformation: The intervention was not only a Keynesian stimulus, but facilitated the move from rural farming areas to the cities where manufacturing was occurring at the time; and facilitated the retraining of the labor force, helping workers acquire the skills necessary for success in an urban manufacturing environment, which were quite different from those that ensured success in a rural, farming environment. It was, in this sense, an example of a successful industrial policy.

There are clear parallels to the situation today in that a significant fraction of the workforce may not have the skills required to succeed in the age of AI.

Transitional Efficiency Wage Theory

Efficiency wage arguments may also slow down the transition to a new equilibrium after technological progress. For example, if worker morale depends on last period's wages, it may be difficult to reduce wages to the market-clearing level after a labor-saving innovation, and unemployment may persist for a long time.[28]

5.3. Jobs and Meaning

An interesting debate that is brought up by the potentially widespread destruction of jobs – discussed already in Keynes' essay on *Economic Possibilities of our Grandchildren* – is that jobs provide not only income but also other mental services such as meaning, dignity, and

[28] In the limiting case, employers may simply keep wages fixed to avoid negative morale effects, and unemployment would persist forever – or until some offsetting shock occurs.

fulfillment to humans. Whether this is a legacy of our past, and whether individuals could find meaning in other forms of activities, mental or physical, is a matter of philosophical debate.

If workers derive a separate benefit from work in the form of meaning, then job subsidies are a better way of ensuring that technological advances are welfare enhancing than simply providing lump sum grants (a universal basic income), as some are suggesting in response to the inequalities created by AI.

This discussion is, of course, a departure from the usual neoclassical formulation, where work only enters negatively into individual's well-being. There are some that claim that individuals' deriving dignity and meaning from work is an artifact of a world with labor scarcity. In a workerless AI world, individuals will have to get obtain their identity and dignity elsewhere, e.g. through spiritual or cultural values. The fact that most humans can find a meaningful life after retirement suggests that there are good substitutes for jobs in providing meaning.

6. Longer-Term Perspectives: AI and the Return of Malthus?

There is a final point that is worth discussing in a paper on the implications of artificial intelligence for inequality and that relates to a somewhat longer-term perspective. Currently, artificial intelligence is at the stage where it strictly dominates human intelligence in a number of specific areas, for instance playing chess or Go, identifying patterns in x-rays, driving, etc. This is commonly termed *narrow* artificial intelligence. By contrast, humans, are able to apply their intelligence across a wide range of domains. This capacity is termed *general* intelligence.

If AI reaches and surpasses human levels of general intelligence, a set of radically different considerations apply. Some techno-optimists predict the advent of general artificial intelligence for as early as 2029 (see Kurzweil, 2005), although the median estimate in the AI expert community is around 2040 to 2050, with most AI experts assigning a 90% probability to human-level general artificial intelligence arising within the current century (see Bostrom, 2014). A minority believes that general artificial intelligence will never arrive. However, if human-level artificial general intelligence is reached, there is broad agreement that AI would soon after become super-intelligent, i.e. more intelligent than humans, since technological progress would likely accelerate, aided by the intelligent machines. Given these predictions, we have to think seriously about the implications of artificial general intelligence for humanity and, in the context of this paper, for what it implies for our economy as well as for inequality.

Assuming that our social and economic system will be maintained upon the advent of artificial general intelligence and superintelligence,[29] there are two main scenarios. One scenario is that

[29] Researchers who work on the topic of AI safety point out that there is also a risk of doom scenarios in which a sufficiently advanced artificial intelligence eradicates humanity because humans stand in the

man and machine will merge, i.e. that humans will "enhance" themselves with ever more advanced technology so that their physical and mental capabilities are increasingly determined by the state of the art in technology and AI rather than by traditional human biology (see e.g. Kurzweil, 2005). The second scenario is that artificially intelligent entities will develop separately from humans, with their own objectives and behavior (see e.g. Bostrom, 2013; Tegmark, 2017). As we will argue below, it is plausible that the two scenarios might differ only in the short run.

First Scenario: Human Enhancement and Inequality

The scenario that humans will enhance themselves with machines may lead to massive increases in human inequality, unless policymakers recognize the threat and take steps to equalize access to human enhancement technologies.[30] Human intelligence is currently distributed within a fairly narrow range compared to the distance between the intelligence of humans and that of the next-closest species. If intelligence becomes a matter of ability-to-pay, it is conceivable that the wealthiest (enhanced) humans will become orders of magnitude more productive—"more intelligent"—than the unenhanced, leaving the majority of the population further and further behind. In fact, if intelligence enhancement becomes possible, then – unless pre-emptive actions are taken – it is difficult to imagine how to avoid such a dynamic. For those who can afford it, the incentive to purchase enhancements is great, especially since they are in competition with other wealthy humans who may otherwise leapfrog them. This is even more so in an economy which is, or is perceived to be, a winner-take-all economy and/or in which well-being is based on relative income. Those who cannot afford the latest technology will have to rely on what is in the public domain, and if the pace of innovation increases, the gap between the best technology and what is publicly available will increase.

A useful analogy is to compare human enhancement technology to healthcare – technology to *maintain* rather than *enhance* the human body. Different countries have chosen significantly different models for how to provide access to healthcare, with some regarding it as a basic human right and others allocating it more according to ability to pay. In the US, for example, the expected life spans of the poor and the wealthy have diverged significantly in recent decades, in part because of unequal access to healthcare and ever more costly new technologies that are only available to those who can pay. The differences are even starker if we look at humanity across nations, with the expected life span in the richest countries being

---

way of its goals. See e.g. Bostrom (2013) who elaborates on this using the example of a "paperclip maximizer" – an AI that has been programmed to produce as many paperclips as possible, without regard for other human goals, and who realizes that humans contain valuable raw materials that should better be transformed into paperclips.

[30] In many respects, the issues are parallel to those associated with performance enhancing drugs. In sports, these have been strictly regulated, but in other arenas, they have not.

two thirds longer than in the least developed countries (see e.g. UN, 2015). Like with healthcare, it is conceivable that different societies will make significantly different choices about access to human enhancement technologies.

Once the wealthiest enhanced humans have separated sufficiently far from the unenhanced, they can effectively be considered as a separate species of artificially intelligent agents. To emphasize the difference in productivities, Yuval Harari (2016) has dubbed the two classes that may result "the gods" and "the useless." In that case, the long-run implications of our first scenario coincide with the second scenario.

Second Scenario: Artificially Intelligent Agents and the Return of Malthus

We thus turn to the scenario that artificially intelligent entities develop separately from regular (or unenhanced) humans. One of the likely characteristics of any sufficiently intelligent entity – no matter what final objectives are programmed into it by evolution or by its creator – is that it will act by pursuing intermediate objectives or "basic drives" that are instrumental for any final objective (Omohundro, 2008). These intermediate objectives include self-preservation, self-improvement and resource accumulation, which all make it likelier and easier for the entity to achieve its final objectives.

It may be worthwhile pursuing the logic of what happens if humans do not or cannot assert ownership rights over artificially intelligent or super-intelligent entities.[31] That would imply that sufficiently advanced AI is likely to operate autonomously.

To describe the resulting economic system, Korinek (2017) assumes that there are two types of entities, unenhanced humans and AI entities, which are in a Malthusian race and differ – potentially starkly – in how they are affected by technological progress. At the heart of Malthusian models is the notion that survival and reproduction requires resources, which are potentially scarce.[32] Formally, traditional Malthusian models capture this by describing how limited factor supplies interact with two related sets of technologies, a production and a consumption/reproduction technology: First, humans supply the factor labor, which is used in a

[31] If humans and artificially intelligent entities are somewhat close in their levels of intelligence, it may still be possible for humans to assert ownership rights over the AI – in fact, throughout the history of mankind, those determining and exerting property rights have not always been the most intelligent. For example, humans could still threaten to turn off or destroy the computers on which AI entities are running. However, if the gap between humans and super-intelligent AI entities grows too large, it may be impossible for humans to continue to exert control, just like a two-year old would not be able to effectively exert property rights over adults.

[32] If AI directs its enhanced capabilities at binding resource constraints, it is conceivable that each such constraint might successively be lifted, just as we seem to have avoided the constraints that might have been imposed by the limited supply of fossil fuels. At present, humans consume only a small fraction – about 0.1% – of the energy that earth receives from the sun. By harvesting energy from beyond earth, even greater supplies of energy would be available. Astrophysicists such as Tegmark (2017) note that the ultimate resource constraint on super-intelligent AI will be the availability of energy (or, equivalently, matter, since $E=mc^2$) accessible from within our event horizon.

production technology to generate consumption goods. Secondly, a consumption/reproduction technology converts consumption goods into the survival and reproduction of humans, determining the future supply of the factor labor.

Throughout human history, Malthusian dynamics, in which scarce consumption goods limited the survival and reproduction of humans, provided a good description of the state of humanity, roughly until when Malthus (1798) published his *Essay on the Principle of Population* to describe the resulting Iron Law of Population. Over the past two centuries, humanity, at least in advanced countries, was lucky to escape its Malthusian constraints: capital accumulation and rapid labor-augmenting technological progress generated by the Industrial Revolution meant that our technology to produce consumption goods was constantly ahead of the consumption goods required to guarantee our physical survival. Moreover, human choices to limit physical reproduction meant that the gains of greater productivity were only partly dissipated in increased population. However, this state of affairs is not guaranteed to last forever.

Korinek (2017) compares the production and consumption/reproduction technologies of humans and AI entities and observes that they differ starkly: On the production side, the factor human labor is quickly losing ground to the labor provided by AI entities, captured by the notion of *worker-replacing technological progress* that we introduced earlier. In other words, AI entities are becoming more and more efficient in the production of output compared to humans. On the consumption/reproduction side, the human technology to convert consumption goods such as food and housing into future humans has experienced relatively little technological change – the basic biology of unenhanced humans is slow to change. By contrast, the reproduction technology of AI entities – to convert AI consumption goods such as energy, silicon, aluminum into future AI – is subject to exponential progress, as described e.g. by Moore's Law and its successors, which postulate that computing power per dollar (i.e. per unit of "AI consumption good") doubles roughly every two years.[33]

Taken together, these two dynamics imply – unsurprisingly – that humans will lose the Malthusian race in the long run, unless counteracting steps are taken, to which we will turn shortly. In the following paragraphs we trace out what this might entail and how we might respond to it. (Fully following the discussion requires a certain suspension of disbelief. However, we should begin by recognizing that machines can already engage in a large variety of

[33] The original version of Moore's Law, articulated by the co-founder of Intel, Gordon Moore (1965), stated that the number of components that can be fit on an integrated circuit (IC) would double every year. Moore revised his estimate to every two years in 1975. In recent years, companies such as Intel have predicted that the literal version of Moore's Law may come to an end over the coming decade, as the design of traditional single-core ICs has reached its physical limits. However, the introduction of multi-dimensional ICs, or multi-core processors and other specialized chips for parallel processing etc. implies that a broader version of Moore's Law, expressed in terms of computing power per dollar, is likely to continue for several decades to come. Quantum computing may extend this time span even further into the future.

economic transactions – trading financial securities, placing orders, making payments, etc. It is not a stretch of the mind to assume that they could in fact engage in all of what we now view as economic activities. In fact, if an outside observer from a different planet were to witness the interactions among the various intelligent entities on earth, it might not be clear to her if, for example, artificially intelligent entities such as Apple or Google control what we humans do [via a plethora of control devices called smartphones that we carry with us] or whether we intelligent humans control what entities such as Apple and Google do. See also the discussion in Turing (1950).) The most interesting aspects of the economic analysis concern the transition dynamics and the economic mechanisms through which the Malthusian race plays out.

In the beginning, those lacking the skills that are useful in an AI-dominated world may find that they are increasingly at a disadvantage in competing for scarce resources, and they will see their incomes decline, as we noted earlier. The proliferation of AI entities will at first put only modest price pressure on scarce resources, and most of the scarce factors are of relatively little interest to humans (such as silicon), so humanity as a whole will benefit from the high productivity of AI entities and from large gains from trade. From a human perspective, this will look like AI leading to significant productivity gains in our world. Moreover, any scarce factors that are valuable for the reproduction and improvement of AI, such as human labor skilled in programming, or intellectual property, would experience large gains.

As time goes on, the superior production and consumption technologies of AI entities imply that they will proliferate. Their ever-increasing efficiency units will lead to fierce competition over any non-reproducible factors that are in limited supply, such as land and energy, pushing up the prices of such factors and making them increasingly unaffordable for regular humans, given their limited factor income. It is not hard to imagine an outcome where the AI entities, living for themselves, absorb (i.e. "consume") more and more of our resources.

Eventually, this may force humans to cut back on their consumption to the point where their real income is so low that they decline in numbers. Technologists have described several dystopian ways in which humans could survive for some time – ranging from uploading themselves into a simulated (and more energy-efficient) world[34] to taking drugs that reduce their energy intake. The decline of humanity may not play out in the traditional way described by Malthus – that humans are literally starving – since human fertility is increasingly a matter of choice rather than nutrition. It is sufficient that a growing number of unenhanced humans decide that given the prices they face, they cannot afford sufficient offspring to meet the

[34]See e.g. Hanson (2016). In fact, Aguiar et al. (2017) document that young males with low education have already shifted a considerable part of their time into the cyber world rather than supplying labor to the market economy – at wages that they deem unattractive.

human replacement rate while providing their offspring with the space, education, and prospects that they aspire to.

One question that these observations bring up is whether it might be desirable for humanity to slow down or halt progress in AI beyond a certain point. However, even if such a move were desirable, it may well be technologically infeasible – progress may have to be stopped well short of the point where general artificial intelligence could occur. Furthermore it cannot be ruled out that a graduate student under the radar working in a garage will create the world's first super-human AI.

If progress in AI cannot be halted, our description above suggests mechanisms that may ensure that humans can afford a separate living space and remain viable: because humans start out owning some of the factors that are in limited supply, if they are prohibited from transferring these factors, they could continue to consume them without suffering from their price appreciation. This would create a type of human "reservation" in an AI-dominated world. Humans would likely be tempted to sell their initial factor holdings, for two reasons: First, humans may be less patient than artificially intelligent entities. Secondly, super-intelligent AI entities may earn higher returns on factors and thus willing to pay more for them than other humans. That is why, for the future of humanity, it may be necessary to limit the ability of humans to sell their factor allocations to AI entities. Furthermore, for factors such as energy that correspond to a flow that is used up in consumption, it would be necessary to allocate permanent usage rights to humans. Alternatively, we could provide an equivalent flow income to humans that is adjusted regularly to keep pace with factor prices.[35]

7. Conclusions

The proliferation of AI and other forms of worker-replacing technological change can be unambiguously positive in a 1st-best economy in which individuals are fully insured against any adverse effects of innovation, or if it is coupled with the right form of redistribution. In the absence of such intervention, worker-replacing technological change may not only lead to workers getting a diminishing fraction of national income, but may actually make them worse off in absolute terms.

The scope for redistribution is facilitated by the fact that the changes in factor prices create windfall gains on the complementary factors, which should make it feasible to achieve Pareto improvements. If there are limits on redistribution, the calculus worsens and a Pareto improvement can no longer be ensured. This may lead to resistance from those in society who

[35] All of this assumes that the super-intelligent AI entities don't use their powers in one way or another to abrogate these property rights.

are losing. As a result, it is desirable to use as broad of a set of 2nd-best policies as possible, including changes in intellectual property rights, to maximize the likelihood that AI (or technological progress more generally) generate a Pareto improvement.

AI and other changes in technology necessitate large adjustments, and while individuals and the economy more broadly may be able to adjust to slow changes, this may not be so when the pace is rapid, for example because of capital market imperfections. Indeed, in such situations, outcomes can be Pareto inferior. The more willing society is to support the necessary transition and to provide support to those who are "left behind," the faster the pace of innovation that society can accommodate, and still ensure that the outcomes are Pareto improvements. A society that is not willing to engage in such actions should expect resistance to innovation, with uncertain political and economic consequences.

References

Acemoglu, Daron. 1998. “Why Do New Technologies Complement Skills? Directed Technical Change and Wage Inequality.” *The Quarterly Journal of Economics* 113(4):1055-1089

_____ 2002. “Directed Technical Change.” *Review of Economic Studies* 69(4): 781-809

Acemoglu, Daron and Pascual Restrepo. 2016. “The Race Between Machine and Man: Implications of Technology for Growth, Factor Shares and Employment.” *NBER Working Paper* 22252.

Aghion, Philippe, Benjamin Jones and Charles Jones. 2017. Artificial Intelligence and Economic Growth. NBER Working Paper 23928.

Aguiar, Mark, Mark Bils, Kerwin Kofi Charles and Erik Hurst. 2017. “Leisure Luxuries and the Labor Supply of Young Men.” *NBER Working Paper* No. 23552

Akerlof, George, and Janet Yellen. 1990. “The Fair Wage-Effort Hypothesis and Unemployment.” *The Quarterly Journal of Economics* 105(2): 255-283.

Arrow, Kenneth. 1962. "Economic Welfare and the Allocation of Resources for Invention." In: *The Rate and Direction of Inventive Activity: Economic and Social Factors*, ed. by Richard R. Nelson, 609-26. Princeton, NJ: Princeton University Press.

Baker, Dean, Arjun Jayadev and Joseph E. Stiglitz. 2017. “Innovation, Intellectual Property, and Development: A Better Set of Approaches for the 21st Century,” AccessIBSA: Innovation & Access to Medicines in India, Brazil & South Africa.

Barrat, James. 2013. *Our Final Invention: Artificial Intelligence and the End of the Human Era*. New York: St. Martin’s Press.

Berg, Andrew, Edward F. Buffie, and Luis-Felipe Zanna. 2016. “Robots, Growth, and Inequality.” *Finance & Development* no. 3 (September): 10-13.

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

Dasgupta, Partha and Joseph E. Stiglitz. 1980a. “Uncertainty, Industrial Structure and the Speed of R&D,” *Bell Journal of Economics* 11(1), pp. 1-28.

_____ 1980b. “Industrial Structure and the Nature of Innovative Activity,” *Economic Journal* 90(358), pp. 266-293.

_____ 1987. “Potential Competition, Actual Competition and Economic Welfare,” *European Economic Review* 32, pp. 569-577.

Dávila, Eduardo, and Anton Korinek. 2017. “Pecuniary Externalities in Economies with Financial Frictions.” forthcoming, *The Review of Economic Studies*.

Delli Gatti, Domenico, Mauro Gallegati, Bruce C. Greenwald, Alberto Russo, and Joseph E. Stiglitz. 2012a. “Mobility constraints, productivity trends, and extended crises.” *Journal of Economic Behavior & Organization* 83 (3): 375-393.

_____ 2012b. Sectoral Imbalances and Long-run Crises. In: *The Global Macro Economy and Finance,* ed. by Franklin Allen, Masahiko Aoki, Jean-Paul Fitoussi, Nobuhiro Kiyotaki, Robert Gordon, Joseph E. Stiglitz. International Economic Association Series. London: Palgrave Macmillan.

Dosi, Giovanni and Joseph E. Stiglitz. 2014. “The Role of Intellectual Property Rights in the Development Process, with Some Lessons from Developed Countries: An Introduction,” in: *Intellectual Property Rights: Legal and Economic Challenges for Development*, Mario Cimoli, Giovanni Dosi, Keith E. Maskus, Ruth L. Okediji, Jerome H. Reichman, and Joseph E. Stiglitz (eds.), Oxford, UK and New York: Oxford University Press, pp. 1-53.

Drandakis, Emmanuel and Edmund Phelps. 1965, “A Model of Induced Invention, Growth and Distribution.” *Economic Journal* 76, 823-840.

Fehr, Ernst and Klaus M. Schmidt. 2003. “Theories of Fairness and Reciprocity – Evidence and Economic Applications.” In: Advances in Economics and Econometrics, Econometric Society Monographs, ed. by Mathias Dewatripont, Lars Peter Hansen and Stephen J Turnovsky. Eighth World Congress, Volume 1, pp. 208-257. Cambridge: Cambridge University Press.

Financial Crisis Inquiry Commission (2011). January, 2011. “The Financial Crisis Inquiry Report”. Final Report of the National Commission. Available at http://www.gpoaccess.gov/fcic/fcic.pdf

Frey, Carl Benedikt, and Michael A. Osborne. 2017. “The future of employment: How susceptible are jobs to computerisation?” *Technological Forecasting and Social Change* 114: 254-280.

Geanakoplos, John, and Herakles Polemarchakis. 1986. “Existence, regularity, and constrained suboptimality of competitive allocations when the asset market is incomplete,” In: Uncertainty, information and communication: Essays in honor of KJ Arrow, ed. by W. Heller, R. Starr, and D. Starrett, pp. 65–96. Cambridge, UK: Cambridge University Press

Gordon, Robert. 2016. *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War*. Princeton, NJ: Princeton University Press.

Greenwald, Bruce, and Joseph E. Stiglitz. 1986. “Externalities in Economics with Imperfect Information and Incomplete Markets.” *Quarterly Journal of Economics 101*(2): 229-264.

Groshen, Erica L., Brian C. Moyer, Ana M. Aizcorbe, Ralph Bradley, and David M. Friedman. 2017. "How Government Statistics Adjust for Potential Biases from Quality Change and New Goods in an Age of Digital Technologies: A View from the Trenches." *Journal of Economic Perspectives* 31(2):187–210

Guzman, Martin, and Joseph E. Stiglitz. 2016a. "A Theory of Pseudo-Wealth," In: Contemporary Issues in Macroeconomics: Lessons from The Crisis and Beyond, ed. Joseph E. Stiglitz and Martin Guzman. IEA Conference Volume, No.155-II. Palgrave Macmillan.

_____ 2016b. Pseudo-Wealth and Consumption Fluctuations," NBER Working Paper No. 22838.

Hanson, Robin. 2016. *The Age of Em*. Oxford: Oxford University Press.

Harari, Yuval N. 2017. *Homo Deus: A Brief History of Tomorrow*. New York: Harper.

Hicks, John 1932. *The Theory of Wages*. London: Macmillan.

Kennedy, Charles. 1964. "Induced Bias in Innovation and the Theory of Distribution." *Economic Journal* LXXIV: 541-547.

Keynes, John Maynard. 1931. "Economic Possibilities for our Grandchildren," in *Essays in Persuasion*, San Diego, CA: Harcourt Brace, pp. 358-373.

Korinek, Anton. 2016. "Currency Wars or Efficient Spillovers? A General Theory of International Policy Cooperation." NBER Working Paper 23004.

_____ 2017. "Humans, Artificially Intelligent Agents, and the Return of Malthus." Working paper.

Korinek, Anton and Ding Xuan Ng. 2017. "The Macroeconomics of Superstars." Working paper.

Korinek, Anton, and Joseph E. Stiglitz. 2017. "Artificial Intelligence, Worker-Replacing Technological Change, and Income Distribution." Working Paper.

Kurzweil, Ray. 2005. *The Singularity Is Near: When Humans Transcend Biology*. New York: Viking.

Lipsey, Richard, and Lancaster, Kelvin. 1956. "The General Theory of Second Best," *The Review of Economic Studies* 24(1): 11-32.

Meade, James E. 1955. *Trade and Welfare*. Oxford: Oxford University Press.

Malthus, Thomas Robert. 1798. *An Essay on the Principle of Population*. Project Gutenberg.

Manning, Alan. 2011. "Imperfect Competition in Labour Markets", In: *Handbook of Labor Economics* edited by O. Ashenfelter and D. Card, volume 4. North-Holland: Amsterdam

Moore, Gordon E. 1965. "Cramming more components onto integrated circuits". *Electronics* 38(8): 114:ff.

Mortensen, Dale T. and Christopher A. Pissarides. 1994. “Job Creation and Job Destruction in the Theory of Unemployment.” *The Review of Economic Studies* 61(3):397-415.

_____ 1998. “Technological Progress, Job Creation, and Job Destruction.” *Review of Economic Dynamics* 1(4):733-753

Newbery, David and Joseph E. Stiglitz. 1984. “Pareto Inferior Trade.” *Review of Economic Studies* 51(1):1-12.

Omohundro, Stephen M. 2008. “The Basic AI drives.” In: Artificial General Intelligence 2008: Proceedings of the First AGI Conference, edited by Pei Wang, Ben Goertzel, and Stan Franklin, pp. 483-492. Amsterdam: IOS.

Piketty, Thomas, Emmanuel Saez and Stefanie Stantcheva. 2014. “Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities,” *American Economic Journal: Economic Policy* 6(1), 2014, pp. 230-271.

Restrepo, Pascual. 2015. “Skill Mismatch and Structural Unemployment.” Working paper.

Samuelson, Paul. 1965. “A Theory of Induced Innovations Along Kennedy-Weisacker Lines” *Review of Economics and Statistics* XLVII: 444-464.

Schmitt, John. 2013. *Why Does the Minimum Wage Have No Discernible Effect on Employment?* Washington, DC: Center for Economic and Policy Research.

Shapiro, Carl and Joseph E. Stiglitz. 1984. “Equilibrium Unemployment as a Worker Discipline Device.” *American Economic Review* 74(3):433-444.

Simsek, Alp. 2013. “Speculation and Risk Sharing with New Financial Assets,” *Quarterly Journal of Economics* 128(3), pp.1365-1396.

Solow, Robert. 1987. "We'd better watch out", New York Times Book Review, July 12, page 36.

Stiglitz, Joseph E. 1969. “Distribution of Income and Wealth Among Individuals.” *Econometrica* 37(3): 382-397

_____ 1976. The Efficiency Wage Hypothesis, Surplus Labour and the Distribution of Income in LDCs. Oxford Economic Papers 28: 185–207.

_____ 1987a. “On the Microeconomics of Technical Progress,” Technology Generation in Latin American Manufacturing Industries, Jorge M. Katz (ed.), New York: St. Martin;s Press, pp. 56-77. (Presented to IDB-CEPAL Meetings, Buenos Aires, November 1978.)

_____ 1987b. “Technological Change, Sunk Costs, and Competition,” Brookings Papers on Economic Activity, 3, pp.883-947.

_____ 2006. “Samuelson and the Factor Bias of Technological Change,” Samuelsonian Economics and the Twenty-First Century, M. Szenberg et al, eds., Oxford University Press: New York, pp. 235-251.

_____ 2008. “The Economic Foundations of Intellectual Property,” Duke Law Journal 57(6), pp. 1693-1724.

_____ 2014a. “Intellectual Property Rights, the Pool of Knowledge, and Innovation.” NBER Working Paper 20014.

_____ 2014b. “Unemployment and Innovation,” NBER Working Paper 20670.

_____ 2014c. “Tapping the Brakes: Are Less Active Markets Safer and Better for the Economy?” Presentation at the Federal Reserve Bank of Atlanta 2014 Financial Markets Conference

_____ 2017. “Pareto Efficient Taxation and Expenditures: Pre- and Re-distribution.” NBER Working Paper 23892

Stiglitz, Joseph E., and Bruce Greenwald 2015. *Creating a Learning Society: A New Approach to Growth, Development, and Social Progress*, with Bruce C. Greenwald, New York: Columbia University Press. Reader’s Edition published 2015

Stiglitz, Joseph E. with Nell Abernathy, Adam Hersh, Susan Holmberg and Mike Konczal, 2015. *Rewriting the Rules of the American Economy,* A Roosevelt Institute Book, New York: W.W. Norton

Stiglitz, Joseph E., and Andrew Weiss. 1981. “Credit Rationing in Markets with Imperfect Information.” *The American Economic Review 71*(3):393-410.

Tegmark, Max. 2017. *Life 3.0: Being Human in the Age of Artificial Intelligence*. New York: Knopf.

Turing, Alan M. 1950. “Computing Machinery and Intelligence,” *Mind* 59(236): 433-460.

United Nations Department of Economic and Social Affairs. 2015. "United Nations World Population Prospects: 2015 Revision."

Vinge, Vernor. 1993. “The Coming Technological Singularity: How to Survive in the Post-Human Era.” In: *Proc. Vision 21: interdisciplinary science and engineering in the era of cyberspace*, pp. 11–22. NASA: Lewis Research Center.

Weizacker, Von, C. 1966. “Tentative notes on a two-sector model with induced technical progress,” *Review of Economic Studies* 33: 245–251.

Williams, Heidi. 2010. “Intellectual Property Rights and Innovation: Evidence from the Human Genome.” *NBER Working Paper* 16213.