Preparing for the (Non-Existent?) Future of Work

by Anton Korinek and Megan Juelfs[*,†]

February 3, 2022

Abstract

We analyze how to set up institutions that future-proof our society for a scenario of ever-more-intelligent autonomous machines that substitute for human labor and drive down wages. We start by evaluating recent predictions about such a scenario and analyze how to allocate work and income if it materializes. As the autonomous income produced by machines rises, we find that it is optimal to phase out work, beginning with workers who have low labor productivity and job satisfaction, since they have comparative advantage in enjoying leisure. This is in stark contrast to welfare systems that force individuals with low labor productivity to work. Providing a basic income, whether in the form of benefits or capital ownership, is the only way to avoid mass misery when there are significant wage declines. Moreover, it gives rise to a multiplier effect that benefits workers – by reducing labor supply, a basic income makes labor scarcer and raises wages. Recipients could still work for its own sake if they derive sufficient utility from the amenities provided by their work, such as structure, purpose and meaning. Finally, society will have to grapple with the externalities and internalities of work amenities to ensure that the transition does not reduce welfare: Some work amenities involve public goods, such as social connections, which may justify job subsidies or public work programs until alternative ways of providing these are developed. Moreover, individuals may have a false sense of the utility they derive from work, leading them to work too much or to withdraw from work too soon. This may create an important role for public education and nudges.

[*]Anton Korinek is a David M. Rubenstein Fellow at the Brookings Institution as well as a Professor of Economics and Business Administration at the University of Virginia, and the Economics of AI Lead at the Centre for the Governance of AI (GovAI). Megan Juelfs is the Associate Director of Research Initiatives at the Institute for Business in Society at the Darden School of Business.

[†]The authors would like to acknowledge financial support from the Center for Innovation, Growth and Society of INET and the Institute for Business in Society at Darden and helpful research assistance by Eleanor Giordano and Marcella Cartledge. We thank Daron Acemoglu, Justin Bullock, Risto Uuk and the participants of the Oxford Handbook of AI Governance conference and the UVA Inequality Working Group for their thoughtful comments.

1 Introduction

In modern societies, work is the main way in which most working-age people earn their income and spend their time. However, many are concerned that advances in artificial intelligence and related technologies may substitute for a growing fraction of workers, presenting significant challenges for the future of work. This paper analyzes how to optimally allocate – or phase out – labor and distribute income in a world in which intelligent machines increasingly become substitutes for human labor.

The paper starts by analyzing under what conditions we would have to be concerned whether work has a future. First, we discuss the technological substitutability of labor. Since the beginning of the Industrial Revolution, technological progress has made production more efficient, but human labor has always remained essential, providing it with a bottleneck role that made labor scarcer and scarcer and led to large gains in wages. However, recent advances in automation have substituted for unskilled labor and have led to wage stagnation in that segment of the labor markets. If advanced AI can fully substitute for any type of labor, then labor is no longer essential for production, and this may mark the beginning of the end of the Age of Labor. We summarize current predictions by technologists regarding the substitutability of labor.

If, additionally, labor-substituting technologies become sufficiently cheap so that they can perform the work of humans at a cost lower than human subsistence consumption, then employing human labor becomes a dominated technology. In such a world, humans could no longer survive based on earning competitive market wages.

We analyze how to optimally allocate work and income as a function of the prevailing state of technology in order to maximize social welfare. For given labor productivity and non-labor income, a worker can find herself in one of three different regimes. When the worker's productivity and non-wage income are too low, she perishes. Once her labor productivity is high enough, she works. However, as her non-labor income rises, it is optimal for her to spend more and more time on leisure rather than work. More generally, technological progress that increases non-labor income or technological changes that reduce the marginal product of labor make it less desirable for humans to work.

When we consider an economy with multiple agents who differ in their labor productivity, we find that increases in wealth make it optimal to phase out work, starting with workers who have low labor productivity and job satisfaction, since they have comparative advantage in enjoying leisure.

The ensuing section of the paper discusses the challenges involved in attaining a desirable income distribution in practice. We analyze what aspects of existing social insurance system need to be reformed to provide foundations for an economy that are robust to a scenario in which the future of work is non-existent.

Then we extend our analysis to account for potential non-pecuniary amenities of work, such as identity, meaning and social connections. Such amenities may make it desirable for some to continue work even if the marginal product of their labor goes to zero. However, we also observe that the amenities of work frequently involve large externalities and internalities that are difficult for markets and workers to deal with.

2 The End of Labor

This section lays out the conditions under which our main scenario of concern may arise, that ever-more-intelligent autonomous machines substitute for human labor and ultimately drive down wages. It starts by presenting a technological precondition that labor can be fully substituted for. It then disucsses recent technological developments and summarizes the leading predictions regarding such a scenario. Next we analyze under what condition the technological substitutability of labor leads to its economic redundancy. We also summarizes and evaluate a few of the main objections to predictions about the economic redundancy of labor.

2.1 Technological Substitutability of Labor

Most predictions about the redundancy of labor are based on the premise that the human brain is at its core a computing device that processes information by transforming inputs into outputs. This premise makes it plausible that advances in hardware and software may catapult the computing capabilities of machines to the point where they are superior to the human brain. When combined with sufficiently advanced sensors and actuators, i.e. robots, this would enable them to perform any kind of work that humans will ever be able to perform. We formalize this concern =as follows:

Definition 1 (Technological Substitutability of Labor)**.** Machines can substitute for any type of labor, i.e. labor is no longer essential for production.

At present, artificial intelligence is clearly nowhere near the ability to perform all human work. Humans still have the edge both in simple non-routine manual tasks and in higher-order cognitive tasks such as understanding the world, planning, and crucially, social intelligence. However, over the past decade, advances in deep learning have enabled artificial intelligence to perform a growing number of less-structured and higher-level cognitive tasks. AI can now accomplish many economically useful tasks at levels of accuracy that either already are or soon will be super-human. Transformers such as OpenAI's GPT-3 can produce high-quality text that is frequently indistinguishable from human-written text and even displays creativity. DeepMind's AlphaFold can predict how the sequence of amino acids encoded in our genes will fold into proteins, opening new avenues for drug discovery and bioengineering. Speech recognition and image recognition tasks can already be performed at super-human levels by state-of-the-art AI. And machine learning has also led to rapid advances in robot dexterity, surpassing human levels in many applications. At present, many of these technologies are at the lab stage, with commercial applications that share their benefits with society at large only at the beginning.

In terms of sheer computing power, the world's most advanced computers are already on par or superior to the human brain. One common measure of computing power are floating point operations per second (flops), corresponding to how many arithmetic operations on real numbers a computer can perform per second. Carlsmith (2020) estimates that the computing power of the human brain can be replicated with about 10^15 flops, given the right software. At the time of writing, Fugaku, the world's top supercomputer that we publicly know about, was able to reach a peak performance exceeding 10^18 flops, easily surpassing this estimate, albeit the system was reported to cost more than $1bn. And computing capacity is expected to continue to grow roughly in line with a generalized version of Moore's Law, which predicts a doubling every two years, for the foreseeable future.

The models underlying cutting-edge AI applications are also experiencing rapid progress. Given the described hardware capabilities, the absence of a human-level general AI that we know of suggests that advances in software are somewhat lagging behind advances in hardware. However, the field is experiencing a rapid inflow of talent and funds (AI Index Report, 2021), suggesting that progress will continue unabated.

The futurist Ray Kurzweil predicts that humans will become technologically redundant in 2035. Bostrom (2014) and Grace et al. (2018) conduct surveys of AI experts on when human-level artificial intelligence may be reached. Bostrom reports a median prediction of 2040 in a sample that included many futurists. Grace et al. report that a broad sample of AI researchers assign a 50% chance that humans will be technologically redundant by the early 2060s.

2.2 Economic Redundancy of Labor

Technological substitutability does not necessarily imply economic redundancy of labor since it does not consider how costly it is to replace human work with machines. If substituting for all human labor is technologically possible, but it would cost many times more than human workers at current market wages, then it is not economically efficient to do so.

The cost at which machines can perform a given human job imposes a ceiling, i.e. an upper bound for the competitive market wage of humans performing the job. As technology advances and machines become more and more efficient, this ceiling – and by implication the market wages of humans, will decline. At first, humans will switch from work that is easily automated to work that is more difficult to automate and therefore pays higher wages, as they have been doing for centuries. However, in the past, there were always sufficient jobs left that only humans could perform. This may no longer be the case if labor is technologically substitutable. And as machines continue to become more efficient across the board, wages may fall so low that human labor is no longer worth it, given the subsistence cost necessary to support a human worker. We formalize this concern as follows:

Definition 2 (Strong Economic Redundancy of Labor)**.** Machines are able to perform any economically valuable task cheaper than humans, valued at their subsistence cost.

Observe that technological substitutability (Definition 1) is a pre-condition for strong economic redundancy of labor (Definition 2) but not vice versa. If the concern captured in this definition is realized, humanity would no longer add economic value - humans would require more economic resources than they are able to create, and human labor would be technologically obsolete. humans could no longer survive based on their competitive market wages alone – a stark departure from the way our societies have been organized since the onset of the Industrial Revolution – and we would have to choose between mass misery or providing a basic subsistence income, as we will explore in more detail below. Definitions 1 and 2 reflect the way in which the potential future redundancy of labor is frequently framed in technology circles.

However, an analogous economic outcome may result from a weaker condition: even if there are still some jobs that can only be done by humans (violating Definition 1) or in which humans are more cost-effective than machines (violating Definition 2), if demand for those jobs is insufficient to provide employment to everyone at a living wage, then the market-clearing price of labor may still fall below the subsistence cost of humans. We formalize this as follows:

Definition 2' *(Weak Economic Redundancy of Labor).* Machines will push the market-clearing price of human labor below the subsistence cost of humans.

Observe that Strong Economic Redundancy (Definition 2) implies Weak Economic Redundancy of Labor (Definition 2') but not vice versa – it all depends on the balance of labor demand and supply not on whether every task can be performed more cheaply by machines. Moreover, Weak Economic Redundancy (Definition 2') can be satisfied even if Technological Substitutability of Labor (Definition 1) is violated, once the number of roles for which labor is still essential is sufficiently limited. Furthermore, we note that Weak Economic Redundancy characterizes the economy's equilibrium and therefore depends not only on technology/labor demand but also on labor supply. For example, if an economy in which labor was weakly redundant introduces a more generous welfare system and labor supply declines, it may lift the weak economic redundancy.

Our discussion so far has focused on machine efficiency and the market price of human labor compared to workers' subsistence cost. Advances in technology may well lead to declining nominal consumer prices as production becomes more efficient. Economic redundancy would only be reached if competitive market wages decline faster than the subsistence cost. If machines become ever more efficient – compared to humans – at transforming inputs such as energy and raw materials into economic outputs, this would likely be the case.

2.3 Objections

Human Superiority One common objection to this perspective is that humans will always create new tasks that machines cannot perform because humans are innately superior to machines in certain domains. This belief is held firmly by many. We acknowledge that this is a possibility, but we also observe that there are no physical or economic laws that would suggest the intelligence and dexterity of machines cannot in principle surpass their human counterparts. Human intelligence is subject to significant natural limits, for example because of natural constraints on the size of our brains. At present, these seem difficult to overcome.

It is certainly clear that advances in the intelligence of machines in recent decades have far outstripped advances in human intelligence. If this progress continues, machines will eventually surpass human levels of intelligence. We also note that for machines to perform all economically productive tasks, they do not necessarily need to have anything corresponding to human consciousness, and they do not need to possess metaphysical attributes that many attribute to humans, such as a soul. They just need to be able to perform all tasks that have economic value in an efficient manner.

Historic Extrapolation Any predictions of human redundancy of labor contrast starkly with the economic experience since the Industrial Revolution, which has been characterized by humans being complemented by machines rather than substituted for. This complementarity has made human labor more and more productive and has raised wages and human living standards by an order of magnitude, ushering in an "Age of Labor." During the 19th and early 20th century, machines replaced our brawn so humans could focus on brain-intensive cognitive tasks. Since the beginning of the computer age, machines have replaced dull and repetitive structured information processing tasks, allowing humans to focus on more interesting and varied cognitive tasks that leveraged our multi-faceted human intelligence.

In short, although past technological advances regularly substituted for human labor in some tasks, they always left other tasks in which labor remained essential and could not be replaced. This made labor a bottleneck in the expansion of output and made it increasingly valuable – as reflected in average wages that have risen more than 20-fold since the beginning of the Industrial Revolution. Aghion et al. (2018) describe in an elegant model how the economy can grow along a balanced growth path if a constant fraction of all the tasks in which human labor is employed is automated every period.

However, there are no fundamental laws for this pattern to continue to hold going forward.

Human Demand Another objection is that an economy could not operate without consumer demand, i.e. that humans need to earn wages so that they can afford to consume goods and services and keep the economy going (see e.g. Ford, 2015). This is a fallacy - it is true that all output produced needs to be demanded by someone or something, but it is not necessary for that demand to derive from humans. Conceptually, it is perfectly feasible for a thriving economy to exist in which all output is used solely for investment purposes, i.e. machines producing output to serve as input for machines (see e.g. Korinek, 2019). The economy would have to re-tool, for example by switching from agricultural farms to server farms that meet the demand created by machines, but there are no economic laws that would make this impossible.

Nostalgic Jobs Another objection is that even if such an economy may be technically feasible, human redundancy will be avoided because humans will always prefer to obtain certain services from other humans rather than from machines, for reasons that we may call "nostalgic." For example, humans may prefer not to replace the services of human priests, judges, or lawmakers.

This reasoning that there are certain human-only jobs relies on three assumptions: First, it requires that we can in fact tell the difference. A robot priest who has greater emotional intelligence than humans and has a more comprehensive understanding of the human psyche than a human priest may well be able to play the role of human priest quite perfectly, or intentionally slightly imperfectly so as to not give away that it is a robot. Second, it assumes that humans will still prefer services performed by other humans – even if machines have a superior track record. For example, properly calibrated artificial judges may be able to make more accurate and humane judgments than humans, leaving behind the noise, discrimination and biases that have plagued our justice system (e.g. Kahneman et al., 2021). Sufficiently advanced autonomous vehicles may kill far fewer people on our roads than human drivers. It seems brutish to insist on jobs being performed by humans in a sub-standard and inefficient way if machines can perform them better and cheaper. Third, it requires that humans earn sufficient income to spend on human services to support human jobs, i.e. that the human share of income (consisting of both their labor and capital income) is sufficiently large.

Comparative Advantage The concept of strong economic redundancy defined above corresponds to machines having absolute advantage in all tasks that can be performed by human labor. By contrast, the theory of international trade tells us that what matters for gainful exchange is comparative advantage, not absolute advantage – more developed nations that are technologically superior in the production of any good can still engage in gainful trade with less developed nations because they export the goods in which they have comparative advantage and import what their trading partner has comparative advantage in. Some argue that comparative advantage should still hold when humans interact with intelligent machines. However, in trade theory, countries are assumed to possess exogenous endowments of factors such as labor that they deploy to their most efficient use – even if these factors earn a pittance. Trade theory does not usually consider whether it is actually cost-effective to maintain these factors. In our setting, by contrast, factors are costly to maintain and producers can choose which technology to pick – under strong economic redundancy, labor is simply a dominated technology that will not be used.

3 Optimally Allocating Work and Income

This section lays out an economic framework to analyze how to optimally allocate work and income. To do so we characterize the first-best outcome of the economy for a given level of technology. The first best describes how to allocate scarce resources in order to maximize welfare and reflects economic allocations under idealized circumstances – it assumes that we can start from an institutional blank slate, i.e. that we can design institutions or adjust existing institutions so as to achieve the described allocation of work and income. This implies that we do not restrict our thinking to conform to the institutional status quo. In short, the first best describes the best possible allocation that an economy could aim for, given the resource constraint, i.e. given the law of scarcity.

We start out by considering the case of a single economic agent – an individual who has a given level of labor productivity and receives a certain amount of non-labor income, e.g. a capital income or a benefit. At first we abstract from any non-monetary amenities that come with work and may affect the agent's welfare such as meaning or social connections – we describe work simply as a transaction that gives up valuable leisure time to produce output. This allows us to illustrate the fundamental forces at work in an intuitive figure. Next we examine the case of multiple individuals who differ in their labor productivity, and we analyze how work and consumption should be allocated between them in an idealized first-best world. Then we extend our framework to incorporate non-monetary amenities and analyze how this should affect the allocation of work and income.

3.1 Work and Income for an Individual Agent

Consider an individual consumer-worker who values consumption c and leisure $1-\ell$, with preferences given by

$$U(c,\ell) = u(c-c_0) + v(1-\ell) \qquad \text{for } c \geq c_0 \text{ and } -\infty \text{ otherwise} \tag{1}$$

where ℓ captures the fraction of time worked, and $u(\cdot)$ and $v(\cdot)$ are both strictly increasing, concave, and satisfy the Inada conditions. We assume that the agent faces a subsistence level of consumption c_0 that is required for her survival. When $c < c_0$, the agent perishes, captured by a utility level of $-\infty$. This is not a significant constraint in today's advanced economies, but it is important to include for some of the adverse scenarios that we are investigating. Observe that our setup reflects that the disutility of the first marginal unit of labor supply at $\ell = 0$ is positive, given by $v'(1) > 0$, because it requires giving up valuable leisure time. We believe that this is an important feature of preferences that is missing in some neoclassical models that impose an Inada condition on the disutility of labor at $\ell = 0$ for analytic simplicity, baking in that agents always supply a positive amount of labor, even if they obtain vanishingly small wages.

The agent's labor is converted into consumption goods according to the production function $y = f(\ell) = w\ell$, where w reflects the agent's labor productivity. The units of

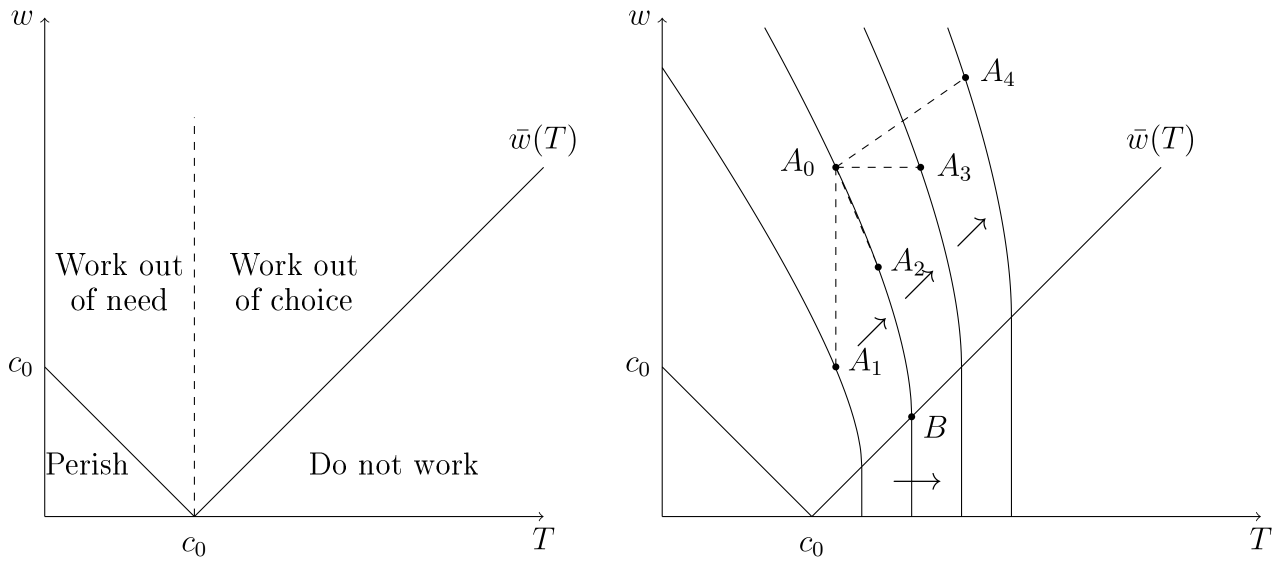

Figure 1: Regions for labor provision and iso-welfare curves

ℓ are chosen such that $\ell = 1$ corresponds to the maximum amount of work possible; therefore w reflects the amount of output produced if the agent engages in maximum work. In a competitive market economy, the labor productivity would correspond to the agent's wage, but our analysis here is focused on the first-best and is more general than a specific market structure.

We assume that the agent also obtains non-labor income in the amount of T, which can be interpreted as income from other factors such as autonomous machines, land or capital, or as a transfer or benefit payment, and which we take as exogenous. Together, this implies a resource constraint that reflects that the agent's consumption needs to be covered by the sum of her labor and non-labor income,

$$c = w\ell + T$$

The first-best in this economy maximizes the agent's utility subject to the resource constraint or, combining the two,

$$\max_{\ell \geq 0} U\left(w\ell + T, \ell\right) \tag{2}$$

Figure 1.a illustrates that the solution to the agent's optimization problem can be decomposed into three different regions as a function of her labor productivity w and the non-labor income T. The figure covers the whole range of possible economic scenarios that an individual agent may face as a result of the increasing automation of labor.

1. (Perish) The triangle at the bottom left of the figure is region 1., which reflects the most dystopian scenario: it occurs when the sum of the individual's potential

labor income and non-labor income are insufficient to meet the subsistence level of consumption. In this region, the individual will perish. This may materialize if the individual's labor productivity diminishes as a result of ever more automation, and if the she does not have sufficient alternative non-labor sources of income to survive. The region is delimited by the minus-45 degree line representing the constraint $T + w = c_0$, which we may call a Malthusian frontier.

2. (Work) To the North-East of the Malthusian frontier is region 2., in which it is optimal for the individual to work. To the left of the dashed vertical line that captures $T = c_0$, it is necessary for the individual to work to cover her subsistence consumption. To the right of the vertical line, she could survive based on the non-labor income T, but it is optimal for her to work as long as her labor productivity is sufficiently high – above the threshold $\bar{w}(T)$, which corresponds to the reservation wage in a competitive economy. $\bar{w}(T)$ is an upward-sloping curve that reflects that a higher non-labor income raises the threshold that makes it worthwhile for an individual to work.

3. (Don't Work) The area to the right of the reservation wage curve is region 3. It captures that the individual's labor productivity is low relative to her non-labor income, making it optimal for her to enjoy her time on leisure instead of working. It would be socially wasteful for the individual to work in this region since the marginal benefit of extra consumption is declining in the level of consumption, and the extra income generated by work would not compensate the individual for her loss of valuable leisure time.

Figure 1.b shows the iso-welfare lines of the agent, which are defined for regions 2. and 3. A given iso-welfare curve depicts all combinations of labor productivity w and non-labor income T for which the individual is indifferent. For example, the individual is indifferent if she moves upward or downward along the indifference curve that goes through point A_0, as any change in her labor productivity is compensated by an off-setting change in the non-labor income. This is illustrated e.g. by a movement from A_0 to point A_2. The arrows in the figure illustrate the direction in which welfare is increasing.

The indifference curves in Figure 1.b allow us to illustrate how technological changes affect welfare. In region 2., there are two distinct channels through which technological change affects individuals' welfare: changes in labor productivity, captured by vertical movements, and changes in non-labor income, captured by horizontal movements in the figure. Starting from an initial point A_0 in Figure 1.b, we illustrate several possibilities:

- Progress that reduces the labor productivity of the agent unambiguously reduces her utility, as reflected in point A_1. As we discussed above in Section 2, this is a scenario that is frequently discussed by technologists who focus on the labor-displacing properties of new technologies.

- By contrast, technological progress that increases the non-labor income T of the agent unambiguously increase her utility, as illustrated in the movement to point A_3. An increase in T arises most directly if the individual holds factors such as capital or land that become more productive as a result of the technological progress. Alternatively, it could also result from institutional changes that provide transfers to the the agent.

- Predictions of future technological progress frequently involve a combination of the two points discussed so far – a diminished role for labor but greater non-labor output. The first effect reduces utility whereas the second effect increases utility, and the overall impact can be either. In our figure, we have illustrated the knife-edge case in which lower labor productivity is precisely offset by higher non-labor income in point A_2. Whenever the increase in non-labor income is sufficient to offset the decline in labor productivity, technological progress leaves the individual better off, i.e. on a higher indifference curve in the figure.

- The best-case scenario is illustrated in point A_4, in which both the agent's labor productivity and her non-labor income go up, increasing the agent's welfare on both counts.

This following proposition describes the solution to the optimization problem formally:

Proposition 1 (Optimal Labor Provision). *The agent's optimum can be decomposed into the following regions:*

1. *(Perish) If $T+w < c_0$, the agent cannot meet the subsistance level of consumption and perishes.*

2. *(Work) If $T + w \geq c_0$ and the agent's labor productivity is sufficiently high, $w > \bar{w}(T)$, the agent works. The reservation level of productivity $\bar{w}(T)$ is given by*

$$\bar{w}(T) := \frac{v'(1)}{u'(T-c_0)} \text{ if } T > c_0 \text{ and } 0 \text{ otherwise} \tag{3}$$

   The optimum amount of labor is determined by equating the marginal benefit of greater output to the marginal cost of forgoing leisure,

$$wu'(w\ell + T) = v'(1-\ell) \tag{4}$$

   An increase in non-labor income T raises utility, decreases optimal labor supply, and raises the threshold $\bar{w}$ for $T > c_0$. An increase in w increases utility and has an ambiguous impact on the optimum amount of labor ℓ. If $T < c_0$, labor must satisfy $\ell > (c_0 - T)/w > 0$ to guarantee the agent's survival.

3. *(Don't Work) If $T > c_0$ and $w \leq \bar{w}(T)$, it is optimal for the agent not to work.*

Proof. If $T+w < c_0$, meeting the subsistence level is not feasible since $c = w\ell + T < c_0$ even for the maximum amount of labor $\ell = 1$, as reflected in point 1.

Otherwise, the optimality condition for labor is

$$wu'(w\ell + T) \leq v'(1-\ell) \tag{5}$$

When $T \leq c_0$, this equation always holds with equality for an optimal level of ℓ that satisfies $\ell > (c_0 - T)/w > 0$. If $T > c_0$ and condition (5) holds as a strict inequality for $\ell = 0$, the extra income from work is not worth the effort, as reflected by the second part of the reservation wage function, and it is optimal for the agent not to work so $\ell = 0$, proving point 3. Otherwise, the optimality condition holds as an equality and pins down a positive optimal level of labor supply $\ell > 0$.

The effects of increases in T and w on utility can be obtained from differentiating the utility function (2) and applying the envelope theorem. The findings on the threshold and on labor supply result from differentiating (3) and implicitly differentiating (4). □

3.2 Multiple Agents With Heterogeneous Productivity

Let us now extend our analysis to an economy in which there are I different types of consumer-workers with utility function (1) that are indexed by $i = 1, ...I$ and differ in their labor productivity w^i. For simplicity we assume that there is a unit mass of each type of agent. A planner who assigns welfare weight θ^i to each agent type i values social welfare according to the function

$$\sum_i \theta^i U\left(c^i, \ell^i\right) = \sum_i \theta^i \left[u\left(c^i - c_0\right) + v\left(1 - \ell^i\right)\right]$$

Given the varying productivity levels among agents, the aggregate amount of effective labor in the economy is given by $\sum_i w^i \ell^i$. For simplicity, we continue to assume a production function that is linear in the aggregate amount of effective labor so

$$Y = F\left(\sum_i w^i \ell^i\right) = \sum_i w^i \ell^i + T$$

where T is an exogenous amount of output produced by automated production processes that do not involve labor. The linear specification of the production function simplifies our analysis but can easily be generalized. (w^i can be interpreted as the marginal product of agent i's labor in equilibrium.)

The planner maximizes welfare subject to the resource constraint

$$\max_{\{c^i, \ell^i \geq 0\}_i} \sum_i \theta^i U\left(c^i, \ell^i\right) \quad \text{s.t.} \quad \sum_i c^i = \sum_i w^i \ell^i + T \tag{6}$$

Condition 1. If $\sum_i w^i + T > Ic_0$, the survival of all agents is feasible.

In words, if the output of all agents working full steam plus the exogenous output T are sufficient to cover the subsistence consumption level of all I agents, it is feasible for everybody to survive. Otherwise, the maximization problem does not have a solution since some agents perish and obtain utility of minus infinity.

In the following, let us assume that the condition is satisfied. The ensuing proposition first lays out the case of an egalitarian planner who assigns equal weight $\theta^i = 1$ to all agents, followed by the case of more general welfare weights.

Proposition 2 (Optimal Labor Allocation With Heterogeneous Productivity)**.** *A. The first-best under an egalitarian planner is characterized as follows:*

1. *There is a reservation level of productivity $\bar{w}$ such that the planner assigns no work $\ell^i = 0$ to all agents with labor productivity $w^i \leq \bar{w}$, and a positive amount of work $\ell^i > 0$ that is increasing in labor productivity w^i to each agent with $w^i > \bar{w}$.*

2. *Increases in the amount of output T produced by automated processes increase the threshold $\bar{w}$ and reduce hours worked for all agents with $w^i > \bar{w}$.*

3. *The planner will assign work to everyone as long as output satisfies*

$$Y < \underline{Y} := I \left(u'\right)^{-1} \left[\frac{v'(1)}{w^{\min}}\right] \tag{7}$$

The planner will completely phase out human labor if

$$T \geq \bar{T} := I \left(u'\right)^{-1} \left[\frac{v'(1)}{w^{\max}}\right] + Ic_0 \tag{8}$$

where $w^{\max} = \max_i\{w^i\}$ is the highest productivity value among all workers.

4. *The distribution of consumption is independent of the distribution of productivity and is given by*

$$c^i = \bar{c} = \frac{Y}{I} \;\forall i \tag{9}$$

B. In the more general case with heterogeneous welfare weights, a greater welfare weight θ^i assigned to agent i implies that the agent is assigned less work and more consumption. Specifically, the productivity threshold $w(\theta^i)$ below which agents do not work is an increasing function of the welfare weight θ^i and labor supply $\ell^i(\theta^i)$ is a declining function of the welfare weight θ^i. Consumption $c(\theta^i)$ satisfies $c(\theta^i) > c_0$ and is an increasing function of the welfare weigth θ^i. Increases in output T from automated processes increase all agents' consumption and reduce the work assigned to all agents. There is a threshold $\bar{T}$ above which human labor is phased out.

Proof. Assigning shadow value λ to the resource constraint in maximization problem (6) and assuming that a subsistence level of consumption for all agents is feasible ($Y > Ic_0$), the planner's first-order optimality conditions are

$$\theta^i u'(\cdot) = \lambda \tag{10}$$

$$\theta^i v'(\cdot) \geq \lambda w^i \tag{11}$$

For part A. with $\theta^i \equiv 1 \forall i$, optimality condition (10) requires that all agents obtain the same level of consumption, which is given by (9) according to the resource constraint, and which pins down $\lambda = u'(\bar{c} - c_0)$. In combination with optimality condition (11), this implies that $\ell^i = 0$ as long as $v'(1) \geq u'(\bar{c} - c_0)\, w^i$ or, equivalently,

$$w^i \leq \bar{w} := \frac{v'(1)}{u'(\bar{c} - c_0)} \tag{12}$$

which takes the same form as 3 under the assumption that $\bar{c} > c_0$. Conversely, for $w^i > \bar{w}$, the optimality condition pins down the amount of labor $\ell^i = 1 - (v')^{-1}[\lambda w^i]$ assigned to agent i, which is increasing in w^i. Since an increase in T raises $\bar{c}$ for a given amount of labor, it lowers λ, raises the threshold $\bar{w}$, and lowers ℓ^i for all agents who work.

The inequality $w^i > \bar{w}$ is satisfied for all agents if it holds for the lowest-productivity value $w^i = w^{\min}$ at given output level Y and consumption level $\bar{c} = Y/I$. Substituting these two values in equation (12) and rearranging delivers the threshold $\underline{Y}$ in equation (7). Conversely, $w^i \leq \bar{w}$ is satisfied for all i if it holds with equality for the highest-productivity agent with $w^i = w^{\max}$ when nobody is working so that $Y = T$. Substituting these two equalities into (12) and rearranging delivers the threshold $\bar{T}$ in equation (8).

For part B., the proof proceeds analogously with the additional complication that each agent's utility is valued differently. □

Figure (2) illustrates the optimal labor provision ℓ^i as a function of an agent's labor productivity w^i . The agent does not work for $w^i \leq \bar{w}$. After this threshold, the optimum amount worked is a concave function of labor productivity. Increases in the output T produced by automated processes shift the reservation wage curve to the right.

Broader Lessons The analysis of agents with differing levels of labor productivity goes to the core of the question of income distribution in a world of increasing automation of labor, but from an idealized perspective. We first observed the condition under which it is feasible for everyone to survive. Given that labor automation promises to greatly increase economic output, this should be an easy condition to meet.

The central insights of the analysis is that in the first-best, the amount that individual agents work is not determined by their individual needs or by their wealth but solely by their labor productivity, i.e. by how much their labor contributes to the production

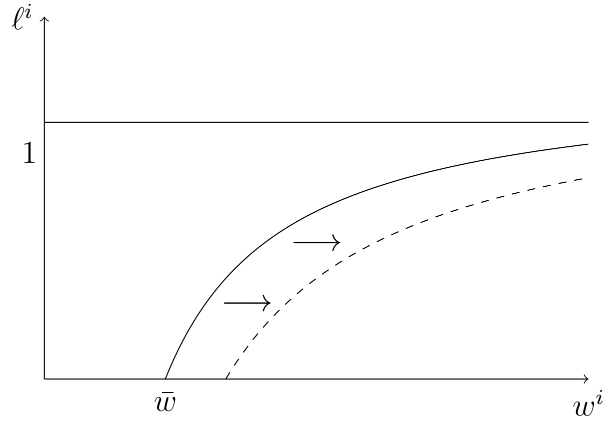

Figure 2: Optimal allocation of labor as a function of labor productivity

of output in the economy. The planner finds it optimal for relatively low-productivity agents to enjoy leisure – they have a comparative advantage in enjoying leisure and can best contribute to overall welfare by enjoying all their time off work. Conversely, the planner assigns work to high-productivity agents, and more so the more productive they are – this is the socially most valuable use of their time. If the wealth created by exogenous automated processes rises, the planner reduces the amount worked for everyone since leisure becomes a more desirable way to spend individuals' time – leisure is a normal good. The planner does so both on the intensive and extensive margin, i.e. by reducing workers' hours and increasing the fraction of the population who do not work. Once the income created by automated processes is sufficiently high, the planner will phase out all human labor – welfare is maximized by letting everybody enjoy their scarce time in the form of leisure.

An interesting observation is that in the described first-best allocation, there is inequality – a utilitarian planner who values everyone equally assigns equal consumption but different levels of work and thus disutility from labor to individuals, depending on their labor productivity. The intuition is that the planner recognizes that individuals with low labor productivity still have the same valuation of leisure as everyone else, and it would be inefficient to make them give up their valuable leisure for work that does not produce sufficient social value.

4 Work Amenities

Work provides a bundle of non-monetary amenities to workers, such as identity, social connections, meaning, and satisfaction, and which directly affect workers' welfare (see e.g. Danaher, 2017). Much of the economic discussion of the future of work focuses on how to provide income to people when they can no longer earn a living wage. However, sociologists, anthropologists, and philosophers have long observed that work also affects workers' welfare through these non-monetary amenities. This becomes especially salient when working-age people lose their jobs (see e.g. Brand, 2015). A comprehensive discussion of economic policy for a work-less future needs to take into account the role of these factors.

We expand our analysis to account for the role of job amenities that affect an individual's utility through channels other than consumption utility or the disutility of providing labor effort. Traditional job amenities include positive factors such as benefits, flexibility, and learning opportunities as well as negative factors like risk exposure, pollution, and arduous working conditions (see e.g. Rosen, 1986). As workers become more prosperous, they also increasingly care about factors such as social connections, identity, meaning, and self-actualization. This can be interpreted through the lens of Maslow's hierarchy of needs as going from deficiency needs to growth needs (Maslow, 1943, 1954).

We capture the amenities that an individual receives from work in the scalar variable a^i, which depends on the amount worked by the individual and enters the utility function as

$$U\left(c^i, \ell^i, a^i\right) \tag{13}$$

where $U(\cdot)$ is quasiconcave and $U_{a^i} \geq 0$. We assume that the work amenities are linear in the amount worked by the individual, $a^i = \alpha^i \ell^i$, where the coefficient $\alpha^i \gtreqless 0$ reflects that work may give rise to either positive amenities or negative disamenities.

Let us define the range of the utility function for zero work as $\mathcal{U}_{\ell=0} = \{u : u = U(c, 0, 0), c > c_0\}$, which indicates all possible utility values that can be obtained by an individual who is meeting the subsistence consumption level but not working.

Lemma 1 (Compensating Differential for Work). *If $U(c^i, \ell^i, a^i) \in \mathcal{U}_{\ell=0}$, then there exists a compensating differential z^i that makes a worker with labor supply $\ell^i > 0$, labor income $w^i \ell^i$, consumption $c^i = w^i \ell^i + T^i$, and amenities $a^i = \alpha^i \ell^i$ indifferent between working and not working,*

$$U\left(T^i + w^i \ell^i, \ell^i, a^i\right) = U\left(T^i + z^i, 0, 0\right)$$

This compensating differential is always positive $z^i > 0$ but can be more or less than the labor income, $z^i \gtreqless w^i \ell^i$ depending on the amenities. If $\alpha^i \leq 0$, then $z^i < w^i \ell^i$. If $\alpha^i > 0$, it is possible that $z^i > w^i \ell^i$. The compensating differential is strictly increasing in the labor productivity w^i of the agent.

Proof. If $U(c^i, \ell^i, a^i) \in \mathcal{U}_{\ell=0}$, by the definition of the range, there exists a z^i such that $U(T^i + z^i, 0, 0) = U(c^i, \ell^i, a^i)$. If $a^i \leq 0$ then observe that $U(c^i, 0, 0) > U(c^i, \ell^i, a^i)$ since utility is strictly decreasing in labor effort and increasing in amenities which are non-positive here. Given that utility is also strictly increasing in consumption, this requires that $z^i < w^i \ell^i$. The compensating differential is increasing in labor productivity since the agent's utility is increasing in w^i. □

The lemma reflects that the total utility impact of work — including the earnings, the disutility of labor, and the individual amenities obtained — can alternatively be derived from a compensating differential of z^i units of consumption goods if the stated condition on the range of the utility function is satisfied. The principle is based on Rosen (1974, 1986)'s equalizing differences, although the lemma here looks at the *total effects* of work rather than solely at job amenities a^i. The overall compensating differential for work is always positive, $z^i > 0$, since not working ($\ell^i = 0$) is in the choice set of the individual (where the choice set is interpreted broadly and may also includes the possibility that the individual does not meet her subsistence consumption level). By revealed preference, if the individual is working, it is because this makes her better off and allows her to earn a utility surplus compared to not working. $z^i > 0$ compensates the individual for this surplus.

The compensating differential that a worker would need to obtain to be made indifferent between working and not working may be more or less than her labor income, $z^i \gtreqless w^i\ell^i$, depending on the value of the amenities associated with the work. In traditional descriptions of the labor market that disregard amenities ($a^i = 0$), the compensating differential is always strictly less than the labor income, $z^i < w^i\ell^i$, since work requires giving up valuable leisure time. In Figure 1.b, for example, the compensating differential for giving up work at, say, point A_0 can be obtained by following the indifference curve to the point where it intersects with the reservation wage curve $\bar{w}$, i.e. to point B, at which the individual ceases to work. The horizontal difference between A_0 and B reflects the transfer necessary to make the individual indifferent between working at A_0 and not working at B. If work involves disamenities to the individual, $a^i < 0$, then the individual needs to receive even less compensation to give up work, and $z^i < w^i\ell^i$ holds a fortiori.

Whenever $z^i < w^i\ell^i$, an individual whose job is displaced can be fully compensated with a benefit that is less than the wage received. Gallup (2022) reports that nearly 85% of employees worldwide and 65% of employees in the US are not engaged or actively disengaged at work, suggesting that they fall into this category. If the job could be automated at zero marginal cost and the surplus could be shared with the individual, overall welfare would increase.

On the other hand, for workers who receive sufficiently positive work amenities, $z^i > w^i\ell^i$ may be the case, i.e. they would need to receive a compensation that is greater than their wage earnings to give up work, since the positive amenities associated with their work more than make up for the loss of valuable leisure time. Mas and Thomas (2022) report that the value of non-wage amenities is positively correlated with wages, suggesting that this phenomenon is more widespread among high earners than low earners. For workers in this category,

Lastly, it is also worth discussing the condition on the range of the utility function of the lemma: intuitively, the condition requires that variations in consumption have sufficient impact on utility to make it possible to compensate individuals for their job amenities. This is an important condition, and it may not always be satisfied: consider an individual who values the meaning derived from her job so greatly that there is no amount of consumption goods that could compensate her for that experience – in that case, a compensating differential for the individual's work does not exist. A stark example in which the condition is violated (which technically cannot be captured by a utility function) is if the individual has lexicographic preferences for certain job amenities. Whenever the individual derives so much utility value from work amenities that she cannot be compensated, it is optimal for her to continue to work, no matter how rich society becomes and no matter how low her labor productivity.

In the following, we assume that the condition of the lemma is satisfied and that agents' utility function has the linearly separable form

$$U\left(c^i, \ell^i, a^i\right) = u\left(c^i - c_0\right) + v\left(1 - \ell^i\right) + x\left(a^i\right)$$

where $x_{a^i} \geq 0$.

We now revisit the question of how to allocate work and income when work is subject to amenities or disamenities, focusing on the case of an egalitarian planner and assuming there is sufficient output for all agents to reach their subsistence level of consumption.

Proposition 3 (Optimal Labor Allocation With Amenities)**.** *A. The first-best under an egalitarian planner with amenities is characterized as follows:*

1. *There is a reservation productivity threshold $\bar{w}\left(\alpha^i\right)$ that is declining in the amenity value of work α^i such that the planner assigns no work $\ell^i = 0$ to all agents with labor productivity $w^i \leq \bar{w}\left(\alpha^i\right)$, and a positive amount of work $\ell^i > 0$ that is increasing in labor productivity w^i to each agent with $w^i > \bar{w}\left(\alpha^i\right)$.*

2. *Increases in the amount of output T produced by automated processes increase the threshold $\bar{w}(\alpha^i) \forall i$ and reduce hours worked for all agents who work.*

3. *The distribution of consumption is independent of the distribution of productivity and the amenity value of work and is given by*

$$c^i = \bar{c} = \frac{Y}{I} \; \forall i \tag{14}$$

4.1 Work Amenities as Externalities

Some amenities are subject to rich externalities. For example, social connections at work entail positive externalities as they arise not from the work of an individual agent but from the collective work of many agents; conversely, the disamenities from commuting become more costly the more workers are commuting. To account for such externalities, we assume that the amenities that a worker obtains from her own work are described by two scalar variables a and $\bar{a}$, where a is an amalgamate of the amenities generated by the individual's labor given by $a = \alpha\ell$ and $\bar{a} = \bar{\alpha}\int_j \ell^j dj$ keeps track of the amenities for the individual worker that are generated by the aggregate amount worked in the economy, where $\alpha, \bar{\alpha} \gtreqless 0$ are parameters. For simplicity, we assume that agent i is atomistic and does not affect the aggregate $\bar{a}$.

We expand the utility function of a consumer-worker to capture $\bar{a}$ in the following linearly separable manner,

$$U\left(c^i, \ell^i, a^i, \bar{a}\right) = u\left(c^i - c_0\right) + v\left(1 - \ell^i\right) + x\left(a^i, \bar{a}\right)$$

where $x_{a^i}, x_{\bar{a}} \geq 0$.

In the following, we focus w.l.o.g. on the case that work amenities are subject to positive externalities from the work of others. (The results flip in case of negative externalities.) To make things tangible, if we focus for example on capturing the social connections generated by work, then a specification with positive $\alpha, \bar{\alpha} > 0$ and a positive cross-derivative such as a multiplicative function $x\left(a, \bar{a}\right) = a\bar{a}$ is suitable.

Proposition 4 (Optimal Labor Choice with Externalities)**.** *If work amenities are subject to positive externalities from the work of others, $\bar{\alpha} > 0$,*

1. *the socially optimal reservation wage is less than what is individually rational,*
2. *the socially optimal choice of labor is greater than what is individually rational,*
3. *the social compensating differential for work is less than that for an individually rational agent,*
4. *individually rational automation decisions are excessive.*

The opposite results hold if work amenities are subject to negative externalities from the work of others, $\bar{\alpha} < 0$.

4.2 Work Amenities as Internalities

There are also indications that some work amenities give rise to internalities for workers, i.e. that individuals deviate from the common standard of perfect rationality when making decisions about their labor choice because they do not internalize the full effect of their choices on their welfare. One interpretation is that the individual who is making the choices is maximizing a different utility function than the one that determines his actual utility. Such internalities can be either positive or negative. For example, a hard-working individual who tells himself that work is a critical component of his happiness even though he would in fact be happier spending more leisure time is experiencing negative internalities from work. Conversely, an individual who lacks structure and meaning and lets his days pass by in loneliness but would in fact be happier holding a job with a regular schedule and regular social interactions would experience positive internalities from work.

We capture the potential for internalities by considering a worker who maximizes a utility function U_0 that disregards the internalities when in fact her welfare is determined by utility function U. For analytical simplicity, we assume that the utility function U_0 completely disregards job amenities and is given by the specification (1), whereas the actual welfare of the individual is determined by utility function (13). This corresponds to the case that all amenities are internalities, and the worker only consider productivity and disutility of labor in her labor choice. (The assumption could easily be generalized at the cost of additional notation.) We find the following results:

Proposition 5 (Optimal Labor Choice with Internalities)**.** *If work amenities are subject to positive internalities, $\alpha > 0$, in the sense that an agent fails to rationally take into account positive amenities from labor, then*

1. *the optimal reservation wage is less than what the agent uses to guide his decisions,*
2. *the optimal choice of labor is greater than what the agent chooses,*
3. *the compensating differential for work is less than what the agent chooses,*
4. *the agent's automation decisions are excessive.*

The opposite results hold if work amenities are subject to negative internalities, $\alpha < 0$.