Nordhaus, William D. 2015. “Are We Approaching an Economic Singularity? Information Technology and the Future of Economic Growth.” NBER Working Paper no. 21547, Cambridge, MA.
Peretto, Pietro F., and John J. Seater. 2013. “Factor-Eliminating Technical Change.” *Journal of Monetary Economics* 60 (4): 459–73.
Romer, Paul M. 1990. “Endogenous Technological Change.” *Journal of Political Economy* 98 (5): S71–102.
Saez, Emmanuel. 2010. “Do Taxpayers Bunch at Kink Points?” *American Economic Journal: Economic Policy* 2 (3): 180–212.
Solomonoff, R. J. 1985. “The Time Scale of Artificial Intelligence: Reflections on Social Effects.” *Human Systems Management* 5:149–53.
Stole, Lars, and Jeffrey Zwiebel. 1996. “Organizational Design and Technology Choice under Intrafirm Bargaining.” 86 (1): 195–222.
Tirole, Jean. 2017. *Economics for the Common Good*. Princeton, NJ: Princeton University Press.
Vinge, Vernor. 1993. “The Coming Technological Singularity: How to Survive in the Post-Human Era.” In *Vision-21: Interdisciplinary Science and Engineering in the Era of Cyberspace*, 11–22. Proceedings of a Symposium Coauthored by the NASA Lewis Research Center and the Ohio Aerospace Institute Held in Westlake, Ohio, Mar. 30–31.
Webb, Michael, Greg Thornton, Sean Legassick, and Mustafa Suleyman. 2017. “What Does Artificial Intelligence Do?” Unpublished manuscript, Stanford University.
Weitzman, Martin L. 1998. “Recombinant Growth.” *Quarterly Journal of Economics* 113:331–60.
Yudkowsky, Eliezer. 2013. “Intelligence Explosion Microeconomics.” Technical Report no. 2013–1, Machine Intelligence Research Institution.
Zeira, Joseph. 1998. “Workers, Machines, and Economic Growth.” *Quarterly Journal of Economics* 113 (4): 1091–117.

Comment Patrick Francois

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

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

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

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

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

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

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

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

The Political Economic Challenges That Machine-Superior Societies Will Face

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

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

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

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

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

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

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

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

The Political Economic Source of Future Human Work

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

Managing the Machines Will Be the Source of Human Work

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

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

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

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

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

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

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

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

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

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

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

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

References

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

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