sanity, humanity and science

real-world economics review

Please click here to support this journal and the WEA

Subscribers: **26,421** **subscribe** RWER Blog ISSN 1755-9472

A journal of the World Economics Association (WEA) 14,468 members, **join**

Sister open access journals: *Economic Thought* and *WEA Commentaries*

back issues

Issue no. 103

31 March 2023

real-world economics review, issue no. 103
subscribe for free

How to Make the Oil Industry Go Bust

Blair Fix [York University, Toronto, Canada]

Abstract
Can the oil industry afford to clean up its mess? If by 'mess' we mean fossil-fuel-induced climate change, the answer is almost certainly 'no'. But what if we look at a more limited cleanup scenario, restricted to the remediation of conventional oil and gas wells? Even then, it seems that the oil industry may already be bankrupt.

As a case study, this paper estimates the solvency of the (conventional) oilpatch in the Canadian province of Alberta. By law, Albertan oil companies are required to pay for well remediation. To date, however, the oil industry has saved little for this expense. Instead, it has assumed that future oil production will pay for existing cleanup liabilities. But will it?

Using cleanup estimates from the Alberta Liabilities Disclosure Project, I estimate the past, present and future solvency of the (conventional) Alberta oil industry. I find that at present, the oilpatch sits on the precipice of insolvency.

The cost of doing business

Environmentalists often argue that if we accounted for the 'true' costs of the fossil-fuel business, the industry would go bankrupt. While I understand this sentiment, I think the idea of 'true' costs is dubious. Costs exist not because they are 'true', but because someone has the power to enforce them. That's why Apple can impose a 30% commission on everything in its app store (Perez, 2020). It has the power to do so.

Back to fossil fuels. Yes, we should make fossil-fuel companies pay for the damage they have wrought. But since doing so will involve a messy power struggle, we need to be pragmatic. On that front, here's a clever option: look for under-enforced legislation that if fully enforced, would make the fossil-fuel business go bankrupt. True, this type of law sounds far-fetched. But it's actually quite common. Here's why.

When fossil-fuel companies extract resources, they are usually required to cleanup their (local) mess. For example, when an oil company drains a well, it's supposed to plug the hole and return the well site to its original condition. Since this cleanup is a legal obligation, you'd think that oil companies — being law-abiding corporate citizens — would have been diligently cleaning up their defunct wells. But for the most part, they let these obligations slide.[1] And so their cleanup liabilities have slowly accumulated.

[1] To avoid their cleanup responsibility, oil companies employ a variety of tactics. The most basic is to simply *delay*. This strategy works because oil companies get to choose when a well is 'done'. Much like when you drink water through a straw, when you pump oil from a well, there are always dregs left over. If you want to delay paying for cleanup, you simply keep the 'straw' open, claiming there are a few more sips of oil to be had.

For years, researcher and activist Regan Boychuk has been watching this crisis unfold (Boychuk, 2010; Boychuk, Anielski, Snow, & Stelfox, 2021). A resident of the oil-rich province of Alberta (Canada), Boychuk is one of the few people paying attention to the oil industry's growing cleanup liabilities. He thinks the procrastination strategy is nearing its end game. If current cleanup liabilities were called in, Boychuk suspects that the oil industry couldn't cover them — not even if it drained every existing well. In other words, Alberta's conventional[2] oil business is *bankrupt*.

Now, you would think that for government regulators, the oil industry's impending bankruptcy would be top of mind. But in Alberta, the regulator is asleep at the wheel. (It's an industry lapdog.) And so Boychuk set out to measure the oil industry's solvency himself. An intrepid muckraker, he acquired the necessary data. But he needed someone to crunch the numbers.

That's where I come in.

Last year, Boychuk asked me if I wanted to join his project, and I immediately said yes. (I grew up in Alberta, so this issue is close to home for me.) This paper is the second entry in our ongoing collaboration. The first piece (Fix, 2022), took a deep dive into Alberta's production of (conventional) oil and gas. In this piece, I'm going to build on that research to analyze the oil patch's solvency.

Speaking of (in)solvency, Boychuk's hunch seems to be correct. According to my estimates, Alberta's (conventional) oil industry now sits on the precipice of bankruptcy. If the government called in the outstanding cleanup liabilities, the oil patch would probably go bust.
And that brings me back to 'true' costs. To bankrupt the oil business, we don't need an expansive new cost framework. By letting its cleanup liabilities pile up, the oil industry has dug its own grave. Now we just need to push it in.

The Alberta oil patch: A brief history

The province of Alberta is a unique place. It sits on the Western edge of the Great Plains — the point where bald prairie gives way to the majestic Rocky Mountains. If we could pick up this landscape and look beneath it, we would see something equally breathtaking. Below Alberta lies a vast web of porous rock containing one of the world's great deposits of oil.

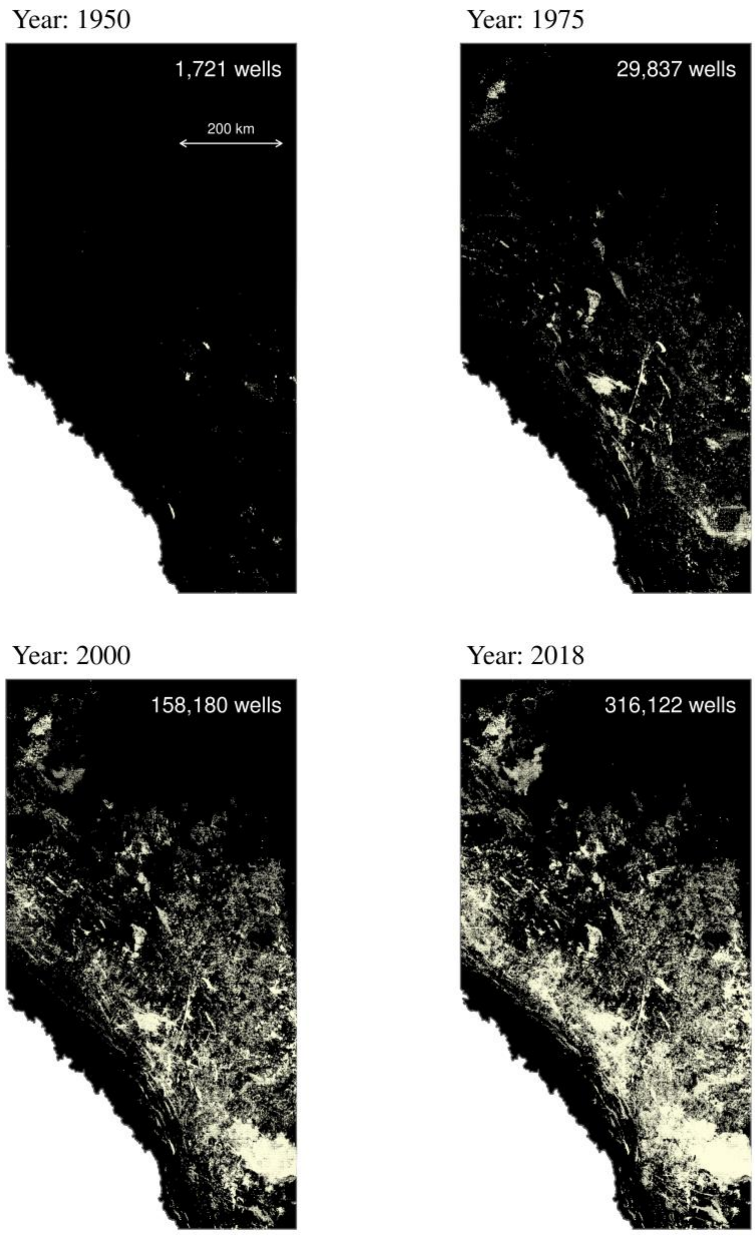
Another tactic is to use creative accounting to get fallow wells off the books. A tried-and-true method is to bundle your defunct wells and sell them to a shell company whose purpose is to do one thing: go bankrupt. Seriously, this happens.

For example, in 2016, Calgary-based oil company Perpetual Energy came up with a clever scheme to get its cleanup liabilities off the books. It took thousands of defunct natural gas wells and lumped them into a bundle of what it called 'mature legacy assets'. Then it created a shell company, Sequoia Resources, which purchased these 'assets' for the price of $1. (Yes, $1 for the whole lot.) A year later, the shell company went belly up, freeing Perpetual of its cleanup debt. In the legal proceedings that followed, we learned that roughly $87 million in cleanup liability was at play (Paperny, Watson, & Frans, 2021).

[2] The term 'conventional' refers to oil and gas that flows from (non-fracked) wells, as opposed to oil from sources like shale deposits or oil sands. Alberta is an interesting case because it has major deposits of conventional oil, but also a monumental reserve of non-conventional bitumen in the Athabasca oil sands. Today, the conventional oil patch is waning, but the exploitation of the oil sands is still accelerating. My analysis here refers only to the 'conventional' oil patch.

Over the last century, Albertans have built a massive web of infrastructure designed to suck this oil and gas from the Earth's depths. Thousands upon thousands of wells have been drilled. If each one of these wells had a bright light on its top, Alberta's nighttime landscape would glow like a bustling metropolis. Figure 1 shows how this satellite view would have evolved.

Figure 1: Illuminating Alberta's oil patch.
This figure imagines what Alberta's nighttime landscape would look like if each oil-and-gas well had a bright light on its top. From 1950 onward, a burgeoning metropolis of wells sprawls across the land. For more details about the data, see the Sources and methods.



From high above, there is a terrible beauty to Alberta's oil-and-gas sprawl. But when we zoom in on the oil business, the picture grows less serene. As you can imagine, drilling for oil is not the cleanest of activities. Figure 2 shows a particularly dirty example — oil leaking from a storage tank east of Calgary.

Figure 2: An oil spill on a well site east of Calgary.
This was the scene in 2016 at an oil well operated by Lexin Resources. The photo was taken by landowner Jim Robins (from Boychuk, 2017).

Because oil companies make a mess when they drill, it's only fair that they clean it up. In Alberta, that's the law. When wells are done producing, they must be plugged, and their surrounding land returned to its original state.

Unfortunately, this responsibility is quite easy to avoid.

In Alberta, the problem has two parts. First, oil companies are responsible for saving for cleanup. (The government doesn't collect funds on their behalf.) Unsurprisingly, most companies save nothing. Second, well cleanup has no definite timeline, meaning companies can delay their responsibilities with little fear of punishment.[3] This strategy works because Alberta's energy regulator is (arguably) an arm of industry. It is 100% industry funded (AER, 2022) and its first board chair was an oil-industry lobbyist.[4]

With the regulator wrapped around its finger, the Alberta oil industry has the power to shirk its cleanup responsibilities … and so it does.

[3] Yes, there are exceptions. For example, the leaking storage tank shown in Figure 2 was operated by Lexin Resources — a company that had accumulated dozens of flagrant environmental violations. In 2017, Lexin was shutdown by the Alberta Energy Regulator (Johnson, 2017).

[4] When the government created the Alberta Energy Regulator in 2013, it appointed Gerry Protti as the board chair (Grandia, 2013). The trouble was, Protti happened to also be the founding president of the Canadian Association of Petroleum Producers — an industry lobby group. He was also the former executive vice-president of Encana, a major Canadian oil company. Critics likened the appointment to putting a fox in charge of the hen house.

Tallying the cleanup bill

After decades of avoiding its cleanup obligations, the Alberta oil industry has racked up an impressive tab. In a moment, I'll show you the bill. But first, let's look at the story behind the data.

Ideally, the Alberta government would maintain a public database to record the oil industry's outstanding cleanup liabilities. No such database exists. Instead, the Alberta Energy Regulator gives cagey estimates based on methods that are unclear. For that reason, Regan Boychuk has always doubted the government's official cleanup numbers. But he had no way of producing a better estimate.

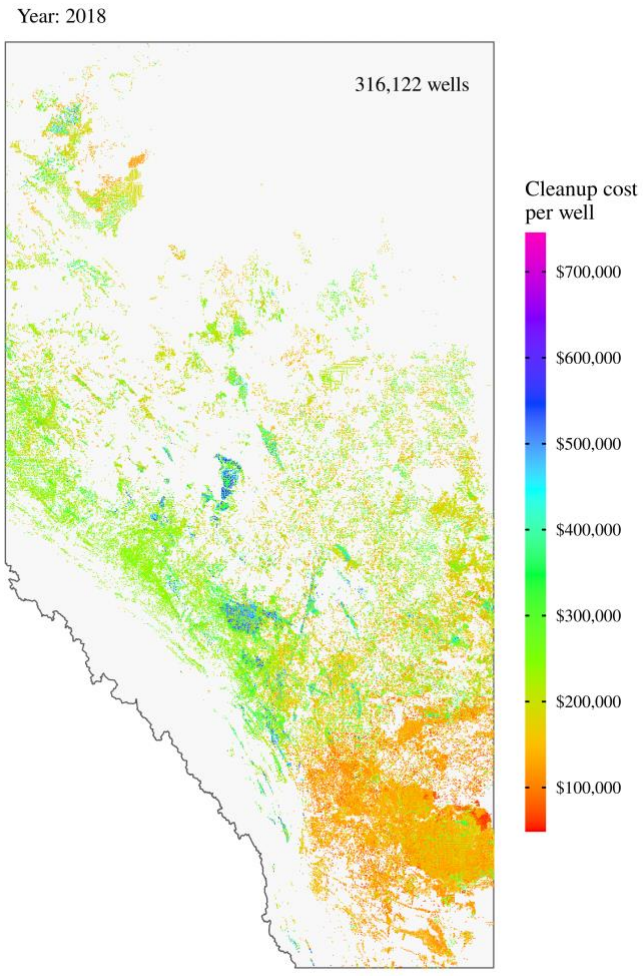
That changed in 2018. During a private presentation in February of that year, Robert Wadsworth — a VP with the Alberta Energy Regulator — revealed some stunning numbers. According to his estimates, the oil industry's cleanup liability was an order of magnitude larger than the regulator's official figure (De Souza, McIntosh, & Bruser, 2018). When news of these estimates made headlines, the government responded by throwing Wadsworth under the bus, claiming his numbers were an 'error in judgment' (Bellefontaine, 2018). His cleanup estimates were never published.

With Wadsworth's data suppressed, the story could have ended there … had Boychuk not been on the case. Luckily, Boychuk filed a freedom-of-information request to ensure that the cleanup estimates saw the light of day. When he got the data, he was floored by its detail. The government database included third-party estimates of a host of cleanup scenarios. And it detailed the features of each Alberta well — features that would affect the cost of remediation. Oddly, however, the government hadn't bothered to apply the cost scenarios to individual wells. Instead, it had come up with some hand-waving averages and called it a day.

And so Boychuk set out to do the job right. Over the next few years, he worked with the Alberta Liabilities Disclosure Project (ALDP) to create rigorous estimates of the Alberta oil industry's cleanup liabilities. They recently published their results in a report called 'The Big Cleanup' (Boychuk et al., 2021). The numbers are astonishing.

Figure 3 illustrates the scale of the problem. Here I've mapped the ALDP cleanup estimates for each well in Alberta (as of 2018). The cost per well ranges from a low of $100,000 to a high of $700,000. If we take these estimates and apply them to the over 300,000 unremediated wells, we have a massive liability. The total bill is between $40 to $80 billion.

Figure 3: A costly legacy.
This map shows the estimated cleanup costs of every conventional oil-and-gas well in Alberta, as of 2018. Each colored point represents a single well. The estimated cleanup cost is shown in color. For more details about the data, see the Sources and methods.



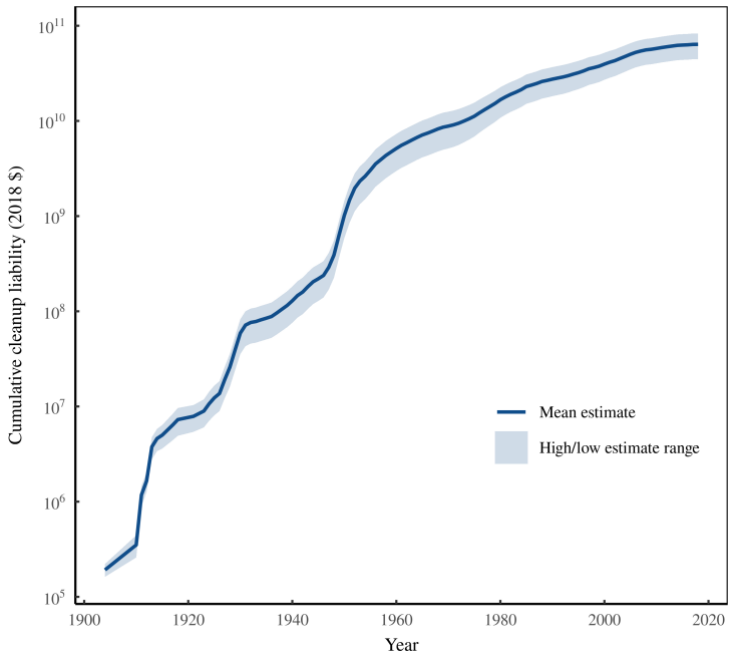
Looking at this billion-dollar liability, you have to wonder — can the oil business afford to pay it off? On that front, what ultimately matters isn't the bill's total size, but how it has (not) been financed.

For example, throughout my lifetime, I will probably spend hundreds of thousands of dollars on groceries.[5] If that bill came due all at once, I'd struggle to pay it. But when administered in small (weekly) doses, I find a way to finance my groceries. I have no other choice.

In Alberta, the oil patch's cumulative cleanup bill is like a grocery tab built up over a lifetime. When you look at the final tally, it's shockingly large. But the bill accrued in small doses. Figure 4 illustrates how it happened. Here, the blue line shows the oil industry's growing cleanup liability, tallied in 2018 dollars. (Note that the vertical axis uses a log scale.)

[5] Suppose you spend $100 per week on food. After 80 years, your grocery tab would be $416,000.

Figure 4: Accumulating a century's worth of cleanup liabilities.
This figure plots the growth of the Alberta oil sector's cumulative cleanup bill, measured in constant 2018 dollars. The blue line shows the mean estimate. The shaded region shows the range for the high/low estimates. Note that the vertical axis uses a logarithmic scale. Moving up the scale, each tick mark indicates that cleanup costs have increased by a factor of 10. For more details about the data, see the Sources and methods.



As of 2018, the oil industry's cleanup liabilities stood somewhere between $40 to $80 billion — a hefty chunk of cash. Surprisingly, most companies haven't saved a dime for this expense. Worse still, their reserves of oil are dwindling, meaning revenues are headed south. So here's the question: how do you finance a growing liability with a diminishing income?

The answer is, you don't. As we'll see, my analysis suggests that the Alberta oil patch is on the verge of insolvency.

To arrive at this end game, we'll take a meandering path. We'll first look under the hood of the oil industry's revenue pump and measure its production of oil and gas. Then we'll use this production data to estimate how much revenue the oil industry raked in throughout its history. As you'll see, it was more than enough money to foot the cleanup bill. Finally, we'll visit real-world Alberta, where the oil business decided to finance cleanup costs using magical thinking. Today, the illusion is about to die.

The revenue pump

Although the logistics of resource extraction can be complex, the oil industry's business model is surprisingly simple. It drills holes in the ground and sells what comes out. In other words, an oil company's revenue depends on two things:

1. the quantity of oil and gas it produces;
2. the price of this oil and gas.

Putting this thinking into an equation, we get the following revenue formula:

$$\text{Revenue} = (\text{Quantity of oil and gas}) \times (\text{oil-and-gas price})$$

To measure revenue, we obviously need to measure both of its components — oil-and-gas production and oil-and-gas prices. And because we're interested in the oil industry's solvency, we need to estimate these components both in the past (which is easy) and in the future (which is harder).

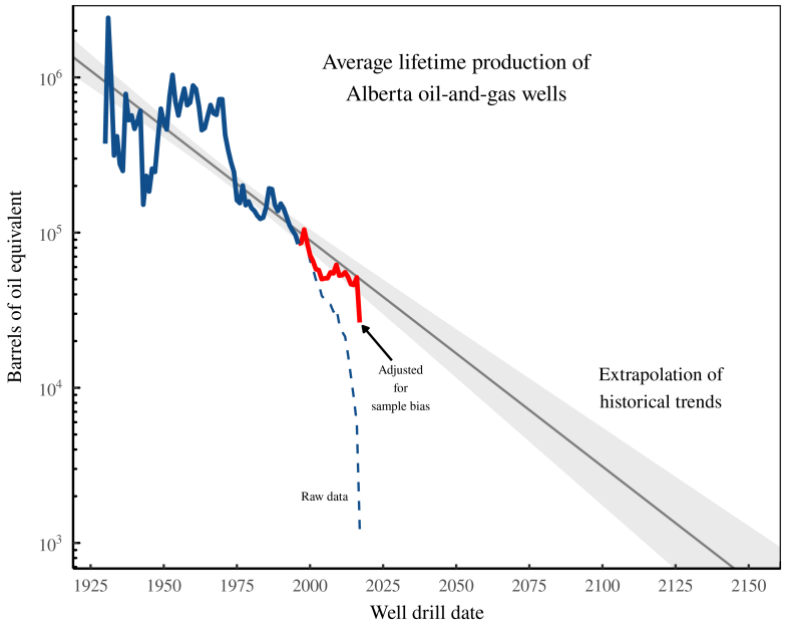
Looking to the future, here's my approach.

Let's start with *prices*. Despite what some analysts claim, no one knows what future oil prices will be … at least not in nominal terms. Yes, we can constrain the price of oil relative to people's income. (For a discussion, see Fix, 2020.) But that doesn't help us predict its nominal price. So when judging the oil industry's future revenue, the best we can do is guess a range of prices that seems plausible. On that front, I'm going to assume oil and gas sells for somewhere between $5 per barrel of oil equivalent (BOE) to $500 per BOE. Note that these values more than cover the historical price range. (See Figure 13 for details.)

Let's move on to the *quantity* of oil and gas produced. Unlike prices (which are a social construct), oil and gas is a non-renewable resource, which means that its extraction follows predictable patterns. As a biological rule, no animal does more work than it must. And so if there is a choice between hard-to-get resources and easy-to-get resources, animals will go for the latter. It's only once the easy stuff is gone that organisms move on to the hard-to-get-stuff.
Humans are no exception to this rule. When it comes to our exploitation of oil and gas, we tap the biggest reserves first. After these reserves are exhausted, we tap the smaller ones. Eventually, there is nothing left but 'puddles'.

Figure 5 illustrates this pattern in Alberta. Here I've plotted the average lifetime production of oil-and-gas wells. Notice how this production declines with time — a nice illustration of our tendency to tap the biggest resources first. During the 1950s, the average well produced the equivalent about one million barrels of oil over its lifespan. Today, that value has fallen by more than an order of magnitude.

Figure 5: The declining lifetime production of Alberta's oil-and-gas wells.
This figure analyzes the lifetime production of Alberta oil-and-gas wells that were, as of 2018, either abandoned or suspended. The blue line (both solid and dashed) shows estimates from raw data. The problem with this raw data is that it becomes biased as we approach the present. By looking only at abandoned/suspended wells, we are selecting wells with the smallest reserve (while those with a larger reserve remain active). The red line shows my attempt to correct for this selection bias. The grey line shows an extrapolation using an exponential fit to historical trends. For details, see the appendix in Fix (2022).

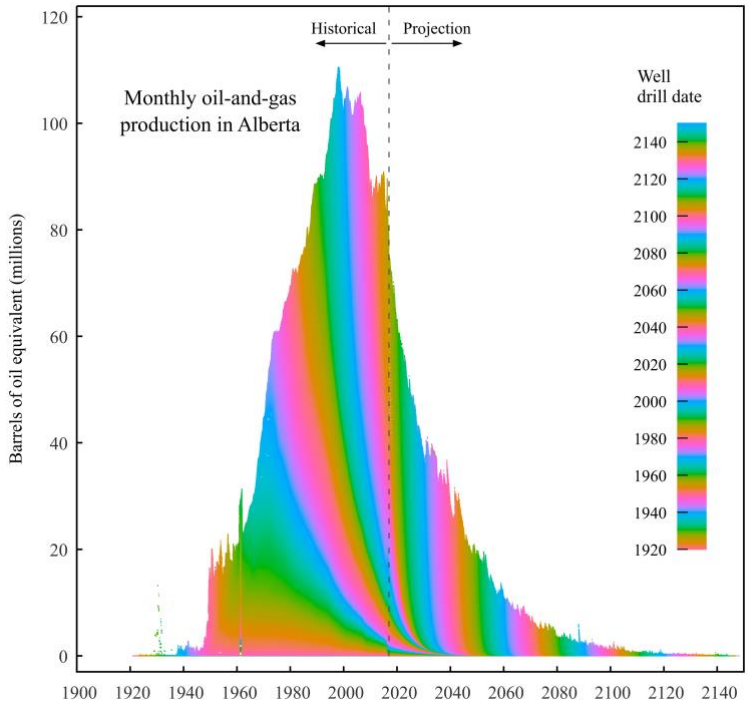


Because oil is a finite resource, we expect that the trend in Figure 5 will continue. After all, once the big reserves are gone, they don't come back. So it's inevitable that, on average, new wells will produce less oil than older ones. Extrapolating this pattern into the future gives us the grey line — a grim prediction of how per-well production will decline over the coming century. By the early 2100s, we predict that the average Alberta well will produce less than 1000 barrels of oil equivalent during its lifetime.

'So what?' you say. Even if per-well production declines, oil companies can counteract the problem by drilling more wells. Yes they can. And historically, that is exactly what oil companies have done. Unfortunately, this strategy solves one problem by creating another. You see, when you drill more wells, you incur more cleanup liabilities … on wells that will themselves produce less oil. And so as you follow the 'drill, baby, drill' strategy, your finances inevitably worsen.
The lesson is this: if we want to understand the oil industry's (in)solvency, we need to model oil-and-gas production (both past and future) at the level of individual wells.

In a previous piece on Alberta oil (Fix, 2022), I developed a model that does exactly that — it resolves Alberta's oil production (past and possible future) down to the individual well. The model is based in part on the production trends shown in Figure 5. I won't review the details here. Instead, I'll cut to the results, shown in Figure 6. Here I've plotted Alberta's monthly production of conventional oil and gas, with color indicating the contribution of individual wells (organized by drill date). The data prior to 2017 is historical. Afterwards, I've used my model to project oil-and-gas production into the future.

Figure 6: Alberta's production of conventional oil and gas — past and possible future.
This figure plots monthly oil-and-gas production in Alberta, resolved to the individual well. The color contours show the production profile as a function of well drill date. Data before 2017 is historical. Data after 2017 is a projection based on an extrapolation of the trends in the empirical data. For more details, see Fix (2022).



Looking at the contours in Figure 6, we can see the well-level trends in action. Over time, the production contours become steeper, indicating that newer wells are exhausted faster than older wells — a pattern that happens almost from day one. But because more wells are drilled, total production does not decline … for a while. Eventually, however, the pace of drilling slows, and production starts to collapse.

If cleanup liabilities are to be funded from future oil-and-gas revenue, this pattern spells disaster. And yet the bankruptcy endgame was not inevitable. As we'll see, the Alberta oil industry had more than enough revenue to cover its cleanup bill. To meet its cleanup obligations, it merely had to save some of its profits.

The Anti Norway

The idea that Alberta oil companies have saved nothing for their cleanup liabilities seems strange … until you understand Alberta's politics. On that front, let's defer to the inimitable Stephen Harper. In 1997, Harper described Canada as "a northern European welfare state in the worst sense of the term" (Mallick, 2015). Unsurprisingly, Harper was both an economist and former employee of Imperial Oil. In 2006, he became Canada's Prime Minister.

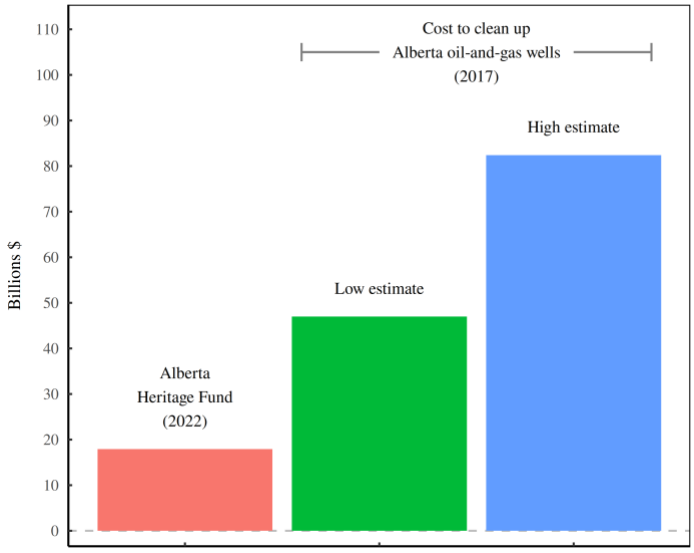
In his disparaging remarks, Harper was railing against 'wasteful' government programs like universal healthcare and public education. But had he described Alberta's regulation of the oil business, he would surely have sung a different tune. When it comes to the oil patch, it would be more accurate to call Alberta a 'North American corporate welfare state'. More specifically, it is the 'anti Norway'.

Like Alberta, Norway has grown rich from its reserves of oil. But unlike Alberta, the Norwegian government chose to *save* most of its oil money. Starting in 1990, it put its oil revenue into a national trust fund. Today, this fund is valued at over $1.2 trillion (NBIM, 2023). The lesson is simple: when we tally the revenue earned from oil and gas, the sum is gargantuan.[6]

Back to Alberta. To be fair, the province does have an oil-and-gas trust fund; it's called the Alberta Heritage Savings Trust Fund.[7] But unlike Norway's colossal trust, Alberta's heritage fund is quite modest. Some history: the heritage fund was created in 1976 by Premier Peter Lougheed with the intention to save 30% of Alberta's oil royalties. A decade later, however, Lougheed's plan was dismantled and trust-fund contributions were abolished. As a result, Alberta's heritage fund is today worth a mere $18 billion (Alberta, 2022).

Now, the phrase 'a mere $18 billion' may sound like an oxymoron. But in this case, the expression is appropriate. According to the government's internal data — data used by Regan Boychuk and the Alberta Liabilities Disclosure Project — the Alberta oil industry is on the hook for a cleanup tab that is between $40 to $80 billion. And unfortunately, the industry hasn't saved a dime. And although the Alberta government *has* saved money, its heritage fund is drowning in the oil-industry's liabilities. Figure 7 shows the water's depth.

Figure 7: The Alberta Heritage fund is smaller than the oil industry's cleanup tab.
The $18-billion heritage fund is what the Alberta government has managed to save from a half-century of oil-and-gas exploitation. Earmarked as 'rainy-day money', the fund is comparably small — less than the oil industry's outstanding cleanup liability. For more details about the data, see the Sources and methods.



[6] In addition to spending their oil revenues differently, Alberta and Norway have very different ways of collecting it. For details, see Wilt (2017).

[7] Alberta does have a cleanup fund of sorts. But it too is laughably small. It's called the 'Orphan Fund', an industry-sponsored plan that currently collects an annual levy of $70 million … for the whole Alberta oil sector (AER, 2021). At that rate, the Orphan Fund will have enough money to cover the oil-sector's present cleanup liability (on the order of $60 billion) in roughly 800 years.

The path untraveled

Clearly, Alberta and Norway took different approaches to managing their oil revenues. Norway used the 'welfare state' model, saving its oil revenue for the people. Alberta opted for the 'corporate welfare' model, sending most of its oil money to the private sector. The difference comes down to politics. On the political spectrum, Alberta is more like Texas than like Norway.

A century ago, however, Alberta was quite different. At the time, it had a budding social credit movement whose goal was to make money creation public by wresting the supply of credit from private banks (see Finkel, 1989 for details). In 1935, the Alberta Social Credit party even won a majority government. But soon after gaining power, the party's politics were co-opted by the flow of oil money. The Social Credit party eventually became a right-wing dynasty — staying in power until 1971.[8]

For arguments sake, let's suppose that history had played out differently. Imagine that Alberta's social credit movement had never died, and that the province remained a bastion of leftist politics. How might this counterfactual Alberta have funded its oil-and-gas cleanup? I think the answer is obvious. The government would have collected money from oil companies and put the funds into a cleanup trust. When wells ran dry, the province would use these funds to pay for remediation.

To investigate how this policy could have worked, let's return to Alberta's real-world accumulation of cleanup liabilities — the trend that I plotted in Figure 4. We're going to compare these growing cleanup costs to the oil industry's cumulative revenue. We'll imagine that *every dollar* of oil-and-gas income is dumped into a simple (interest free) savings account. Then we'll ask: how would this savings account grow with time? And would it cover the industry's growing cleanup liabilities?

To answer these questions, I'll use the oil-and-gas production data shown in Figure 6. (Remember that this data is partly historical and partly a projection.) Here's my method:

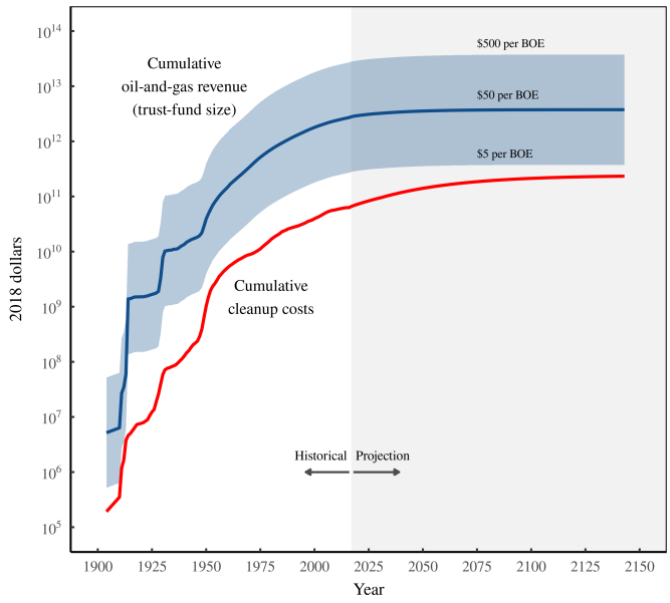
1. To calculate the size of our savings fund in a give year, I first sum the cumulative production of oil and gas up to that year.
2. I then multiply this cumulative production by a hypothetical price. To cover a wide range of scenarios, I'll assume that oil and gas varies between $5 per BOE (barrel of oil equivalent) and $500 per BOE, with a midpoint of $50 per BOE. (Note that these values generously cover the historical range of oil and gas prices. See Figure 13 for details.)

Using this method, Figure 8 shows how the oil industry's finances would have played out. The blue curve and shaded region show the size of the cleanup trust fund (measured in 2018 dollars, and assuming various prices of oil and gas).[9] The red curve shows the growth of cleanup liabilities.

[8] After losing power in 1971, the Social Credit party moved to the fringes. From 1982 onward, it failed to win any seats in the provincial legislature. In 2017, the party was renamed the Pro-Life Alberta Political Association, indicating its sole focus on the issue of abortion.

[9] My trust-fund calculation is what physicists would call a 'Fermi estimate'. By ignoring interest and inflation, I'm simplifying the real world to get a ballpark estimate of how much oil-and-gas revenue Alberta could have collected.

Figure 8: A counterfactual Alberta with a cleanup trust fund.
This figure imagines a counterfactual Alberta in which all oil revenues are dumped into a cleanup trust fund. The blue curves show the growth of this fund for various prices of oil and gas. (To create this curve, I use the oil-and-gas production data from Figure 6.) The red curve shows the growth of cumulative cleanup costs. Data prior to 2017 is based on historical oil-and-gas production. (Oil-and-gas prices are counterfactual.) Data beyond 2017 is based on projected oil-and-gas production. See Figure 6 for details. For more details about the data, see the Sources and methods.



Looking at Figure 8, we can see that our (counterfactual) trust fund grows faster than the oil industry's cleanup costs. That's good. It means that if the Alberta oil industry had planned ahead, it could have easily funded its own cleanup.

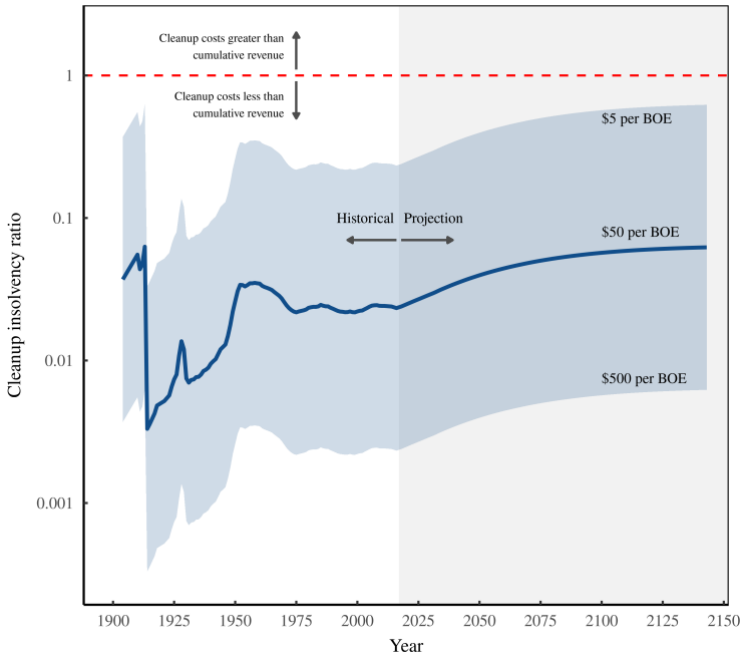
To dive into these finances further, we can calculate something that I call the 'cleanup insolvency ratio'. This is the ratio between the oil industry's cumulative cleanup liability and its cumulative revenue:

$$\text{cleanup insolvency ratio} = \frac{\text{cumulative cleanup liability}}{\text{cumulative oil-and-gas revenue}}$$

The cleanup insolvency ratio measures the oil industry's ability to fund its own cleanup. To avoid bankruptcy, the oil business must keep its insolvency ratio below one.

As Figure 9 illustrates, our counterfactual Alberta meets this criterion easily. Even for the low oil-and-gas price of $5 per BOE, the insolvency ratio remains below one. In other words, the oil industry is solvent for its entire lifespan.

Figure 9: The cleanup insolvency ratio in a counterfactual Alberta.
The cleanup insolvency ratio (blue line and shaded region) measures the oil industry's ability to finance its own cleanup. In this counterfactual scenario, the insolvency ratio divides the oil industry's cumulative cleanup costs by its cumulative revenue. When the insolvency ratio is less than one (i.e. below the dashed red line), the oil industry can afford to fund its own cleanup. Here, I calculate the insolvency ratio for various prices of oil and gas. Note that data prior to 2017 is based on historical oil-and-gas production. (Oil-and-gas prices are counterfactual.) Data beyond 2017 is based on projected oil-and-gas production. See Figure 6 for details. For more details about the data, see the Sources and methods.



With this thought experiment, we've shown that the Alberta oil industry made more than enough money to fund its own cleanup. To avoid the real-world crisis that it now faces, the industry merely needed to save some of its revenue. On that front, the insolvency ratio tells us the required savings rate. For example, if the insolvency ratio is 0.1, we know that cleanup costs represent 10% of cumulative revenue. Flipping things around, it follows that if the industry saved 10% of its revenue, it could finance its own cleanup.

Looking at Figure 9, we can see that the necessary savings rate depends on the price of oil and gas. As an example, suppose that oil and gas sells for a constant $50 per BOE. At that price, Alberta's cleanup insolvency ratio averages 0.04. In other words, the oil business could have funded its own cleanup by saving 4% of its revenue.

Now, Alberta's history shows that when oil companies are left to their own devices, they won't put this money away. Fortunately, we have things called *taxes* that allow governments to collect non-optional payments from citizens and businesses. For example, Canadian consumers pay a 5% federal sales tax on most purchases. Had the Alberta government applied the same tax rate to the oil business (and saved the resulting income), there would likely be no cleanup crisis.
To wrap up this counterfactual journey, my estimates confirm what you probably expected: the Alberta oil business has historically bathed in riches. If the industry had saved a small fraction of its revenue, the cleanup crisis wouldn't exist.

Funding cleanup with magical thinking

Back to real-world Alberta. Instead of choosing the tedious (but effective) method of saving for its looming cleanup, Alberta's oil industry opted for a more inventive approach; it appealed to magic.

Here's a step-by-step guide for how it works:

1. Drill a well.
2. Record the cleanup cost.
3. Put this liability aside and forget about it.
4. Drain the well and distribute the profit to shareholders. Save nothing.
5. Assume that future income will magically cover the cleanup costs.
6. Realize that your well has run dry and there is no future income.
7. Declare bankruptcy and (magically) walk away.

Now, I'm ridiculing this let-the-future-pay-for-cleanup funding model because at its core, it is ridiculous. Oil is a finite resource, which means that in the future, oil revenue will inevitably dwindle. When that happens, it becomes impossible to finance the cleanup liabilities for which you've saved nothing. In other words, the magic of this model is its appeal to *corporate bankruptcy*, which lets you magically wipe liabilities from the books. Over the long term, the let-the-future-pay-for-cleanup model is a catastrophe waiting to happen.

What's interesting, though, is that in the early days of oil exploration, this funding model seems to work. That's because when you continuously discover new oil reserves, your assets (future oil income) grow faster than your cleanup liabilities. So even though you've saved nothing to cover your cleanup costs, your finances look sound. The problem, of course, is that new discoveries eventually dry up. When they do, your finances head south.

Adding up the future

To track the oil industry's finances, we need to quantify the value of its future income. Now, this kind of valuation is something that investors do all the time. But we will *not* do what they do.

Investors start by assuming that today's income stream will continue indefinitely. So if a company averaged a million dollars of annual profit over the last few years, investors assume that the company will continue to receive this income forever. The problem with this assumption is that when you add up a perpetual income stream, you get a value that is infinite. That's no good. To sidestep the infinity, investors 'discount' the future. They look at income that will be received ten years from now and value it less than the same income earned tomorrow. By doing so, the investor converts an infinite stream of revenue into a finite present value.

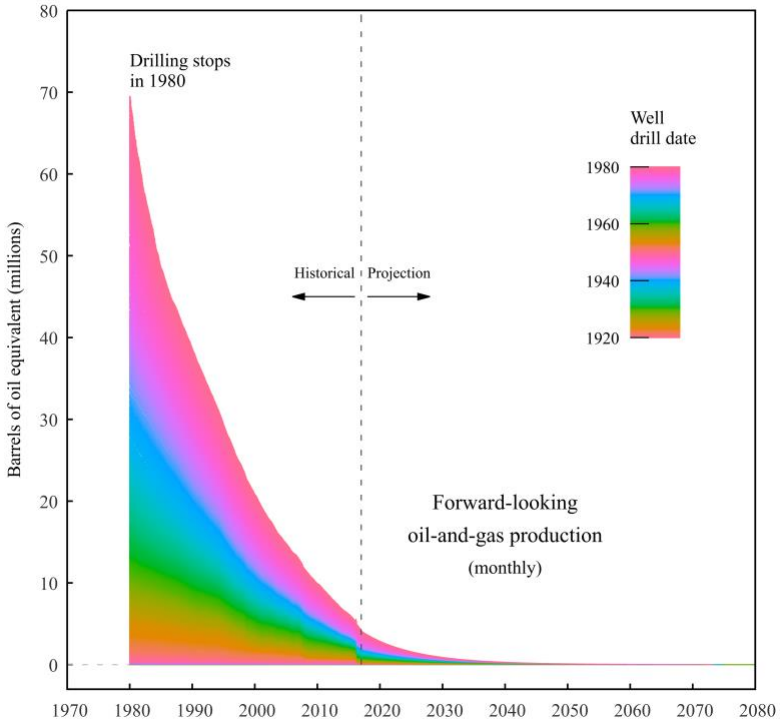
This practice is called 'capitalization', and it is *not* appropriate for valuing the future income of the oil industry.[10] That's because the oil business has an income stream that is decidedly non-perpetual. Oil companies sell a finite, non-renewable resource. So their future income is, by definition, finite. In other words, we (the analysts) don't need to discount the oil industry's revenue stream. Nature does that for us.

[10] Actually, capitalization is not appropriate for *any* scientific valuation. Capitalization is a ritual — part of the ideology of capitalism. For a detailed discussion, see Fix (2021).

Nature's discount rate is part of the dynamics of oil production. As you drain a well, its production tends to decline exponentially with time. This decline means that even if a well was operated forever, its cumulative production of oil will still be finite. (That's how the mathematics of exponential decay work.) So when calculating an oil company's future income, we simply need to observe the rate by which production is declining, and extrapolate this rate into the future. When we sum the resulting production, we're left with a finite quantity of oil — a finite quantity that serves as the oil industry's sole source of income. If we then multiply this quantity of oil by an assumed price, we've estimated the industry's total future income.

To see these exponential dynamics in action, let's turn to Figure 10. Here, I've imagined a counterfactual Alberta in which oil-and-gas drilling stops in 1980. Given this drilling moratorium, we then observe how oil-and-gas extraction would play out (as existing wells are drained). It's a future marked by continuous exponential decline. (Note that data prior to 2017 is historical. Data afterwards is a projection.)

Figure 10: Alberta's monthly production of oil and gas, supposing that drilling stopped in 1980.
This figure imagines a world in which the Alberta oil industry halts drilling in 1980, and then drains its existing wells. The color contour shows the monthly oil-and-gas production that would result. Production data is from the scenario shown in Figure 6. Data prior to 2017 is historical. Data after 2017 is a projection. For more details about the data, see the Sources and methods.



Now, in real-world Alberta, there was no drilling moratorium. Instead, oil companies drilled many more wells, which allowed them to produce much more oil than shown in Figure 10. So when considering the oil industry's future income, shouldn't we include the production from future wells?

When it comes to evaluating cleanup (in)solvency, the answer is *no*. That's because drilling more wells means incurring more cleanup liabilities. So it's not fair to compare the revenue from future wells to today's cleanup liabilities. The bottom line is that if the income from existing wells won't cover your outstanding cleanup costs, then drilling more wells will only make the problem worse.

Back to Figure 10. The area under the production curve represents the oil industry's biophysical assets as of 1980. By my estimates, it amounts to about 12.5 billion barrels of oil equivalent waiting to be extracted. To convert this biophysical quantity into a monetary value, we must assume some price of oil. Supposing that oil and gas sells for $50 per BOE, we find that as of 1980, the Alberta oil industry had about $625 billion of future income sitting in its existing wells.

As it happens, $625 billion was more than enough money to pay for the industry's cleanup liability. In 1980, that liability stood at $17 billion (measured in 2018 dollars). So here we see the initial appeal of the let-the-future-pay-for-cleanup funding model. In 1980, the Alberta oil industry had plenty of big reserves to exploit. So it had lots of future income to cover its existing cleanup liabilities. In short, had the oil business thrown in the towel in 1980, it would've had a cushy landing. It could have drained its existing wells, put a small portion of the resulting revenue towards well cleanup, and then pocketed the rest.

Of course, the Alberta oil industry did not throw in the towel. Instead, it kept drilling. By doing so, it slowly dug its own grave.

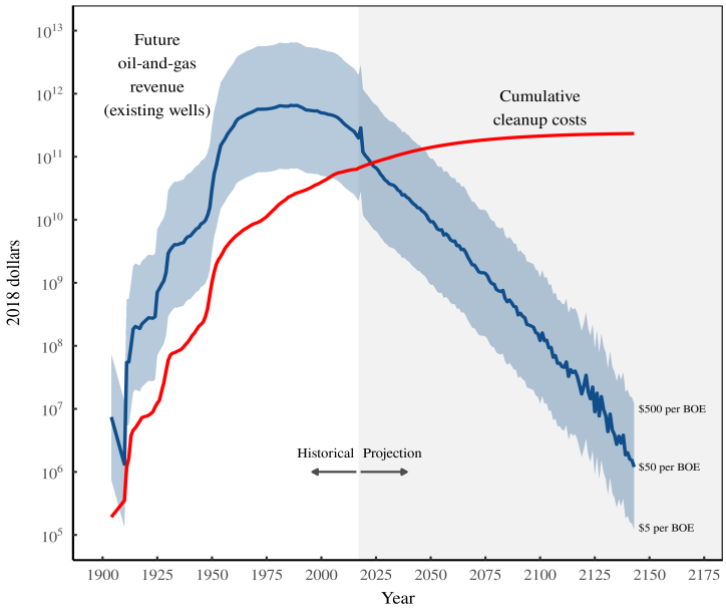
Drilling down to insolvency

We're now ready to look under the hood of the Alberta oil patch's finances. On that front, let's take a peek at Figure 11. What it shows is rather alarming.

Conceptually, this chart is similar to Figure 8, which as a reminder, analyzed the finances of a counterfactual Alberta that *saved* its past oil-and-gas revenue. The big difference with Figure 11 is that it's concerned with the oil industry's *future* income. Why? Because in real-world Alberta, past revenues have all been spent. The only option for funding well cleanup is to hope the future can pay for it.

On that front, this policy actually worked … for a while. In Figure 11, the blue line and shaded region show the Alberta oil industry's future income as a function of time (assuming various prices of oil and gas). Until the 1970s, future income actually grew *faster* than cleanup costs (which are shown in red). This pattern is expected — it's what happens when you discover big new reserves. But what's also expected is that as you exhaust these reserves, your fortunes will reverse. In Alberta, the U-turn took hold during the 1990s. At the time, the drilling pace remained furious, so cleanup liabilities kept piling up. But each new well sucked out less and less oil, meaning future income began to shrink. As a consequence, the oil industry's finances began to sprint south.

Figure 11: Alberta's let-the-future-pay-for-cleanup scheme worked … for a while.
This figure shows how Alberta's let-the-future-pay-for-cleanup funding model is playing out. The red curve shows the growth of cleanup liabilities, which accumulate as new wells are drilled and not cleaned up. The blue curve shows the assets that will pay for this cleanup — the industry's future income (assuming various prices of oil and gas). These estimates are based on the production scenario shown in Figure 6. Data prior to 2017 is based on historical oil-and-gas production. (Oil-and-gas prices are counterfactual.) Data beyond 2017 is based on projected oil-and-gas production. Note that since future income is forward looking, the dividing line between historical and projected revenue is fuzzy. For more details about the data, see the Sources and methods.



When we think through the let-the-future-pay-for-cleanup strategy, we can see that it's a game with only one end. At a certain point, cleanup liabilities become larger than future income. When that happens, the oil industry is sunk.

I suspect that government regulators know how this game works. But to the degree that they think about it, my guess is that regulators treat the oil industry's bankruptcy as a distant problem.
It is not.

Looking at Figure 12, we can see that the oil industry's cleanup costs surpass its future revenue right about now.

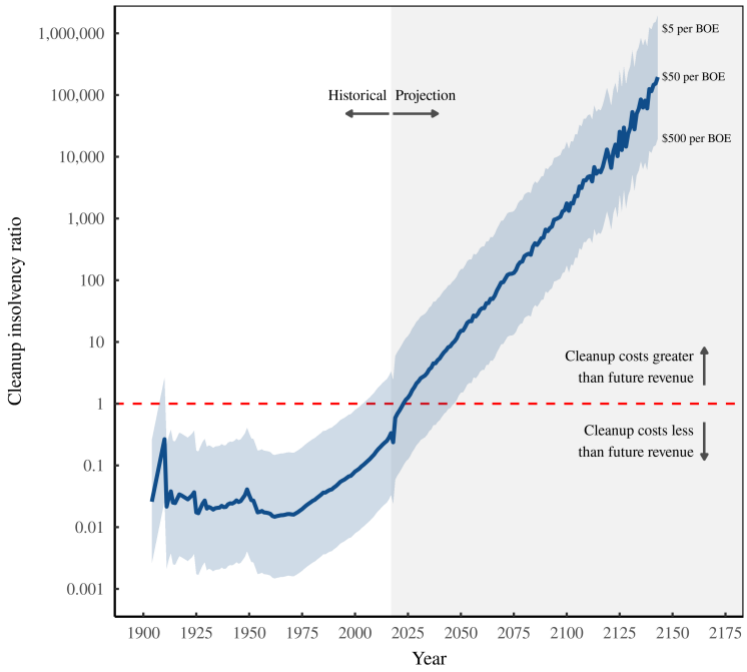
To show this timeline more clearly, let's return to the measure that I call the 'cleanup insolvency ratio'. This ratio takes the oil industry's cumulative cleanup liabilities and divides them by a measure of income. In my counterfactual example (Figure 9), that income came from *past* oil-and-gas revenue. But in real-world Alberta, cleanup costs are supposed to be covered by *future* oil-and-gas revenue. So the cleanup insolvency ratio becomes:

$$\text{cleanup insolvency ratio} = \frac{\text{cumulative cleanup liabilities}}{\text{future oil-and-gas revenue}}$$

As a reminder, the oil industry is solvent when the cleanup insolvency ratio is less than one. When the insolvency ratio surpasses one, the oil industry is formally bankrupt.[11]

Figure 12 shows how the insolvency ratio plays out in real-world Alberta. As anticipated, things start out well. During the early years of exploration, the Alberta oil industry had an insolvency ratio well below one. But after 1970, the insolvency ratio began to creep upward, slowly approaching the threshold for bankruptcy (the dashed red line). The exact timing of this bankruptcy depends on the assumed price of oil and gas. At $500 per BOE, the industry will be bankrupt in a few decades. At $5 per BOE, the industry is already bankrupt, and has been for two decades. And at $50 per BOE, the industry is going bankrupt *now*.

Figure 12: The cleanup insolvency ratio in real-world Alberta.
This figure plots the insolvency ratio for the oil industry in real-world Alberta. This ratio divides cumulative cleanup costs by the oil industry's future income. (I plotted the insolvency ratio's components in Figure 11.) The line and shaded region show the value for different prices of oil and gas. Data prior to 2017 is based on historical oil-and-gas production. (Oil-and-gas prices are counterfactual.) Data beyond 2017 is based on projected oil-and-gas production. Note that since future income is forward looking, the dividing line between historical and projected revenue is fuzzy. For more details about the data, see the Sources and methods.



Because the future price of oil and gas is unknown, we can quibble about the timing of the oil patch's insolvency. The truth is that we'll only know the exact date in hindsight. But by then, it will be too late; the prospect of an industry-funded cleanup will be dead. And so it behooves us to look at the big picture. If the status quo continues, what is certain is that the Alberta oil industry is headed for bankruptcy.

[11] A cleanup insolvency ratio of 1 is the hard limit above which the oil industry cannot possibly be solvent. Above this limit, cleanup costs exceed future revenue. In practice, however, oil companies will be bankrupt well below this threshold. After all, cleanup is but one of many costs that oil revenue must cover. Where this practical bankruptcy occurs depends on the particulars of each company's finances. At any rate, the Alberta oil industry is approaching (or may have already passed) the hard limit of insolvency. So the exact location of the softer limits is irrelevant.

Importantly, we're not talking about 'minor' bankruptcy either. No, the bankruptcy that awaits Alberta's oil patch is so complete that it boggles the mind.[12] On that front, notice that in Figure 12, the vertical axis uses a logarithmic scale, meaning each tick mark indicates a factor of ten. So as the cleanup insolvency ratio heads upward, it does so at an exponential pace. The results are catastrophic. In the not-so-distant future, there will come a day when the oil industry's cleanup liabilities exceed its future income by a *thousand-fold* (or more).

At that point, the oil industry has become a vampire. It sucks profit from society and leaves behind a trail of corpses that it cannot possibly cleanup. Let's hope that long before this day comes, someone finds a stake.

Hastening the inevitable

From the moment the first veins of coal were opened (thousands of years ago), one thing has been certain: the fossil-fuel business would eventually die. But what's always been uncertain is the when and the how. That's because there is no law of nature that tells us how much of a non-renewable resource humans will exploit.

One possibility is that we will harvest fossil fuels to the point of utter exhaustion. Of course, there will always be some scraps left over. But eventually, fossil fuels will become so sparse that we'll put more energy into harvesting them than we receive back. When that happens, the game is up. But although this end game is conceivable, the realities of climate change make it suicidal. There are likely enough fossil fuels left in the ground to render much of the Earth uninhabitable.

Faced with the specter of climate change, another possibility is that technology will come to the rescue. On that front, there is some historical precedent. Humans didn't stop riding horses because we ran out of grass. We stopped because grass-eating horses were replaced by gas-burning cars. So perhaps we'll stop harvesting fossil fuels because renewable energy renders them obsolete. Admittedly, this scenario is comforting. Unfortunately, there's little evidence that it is actually happening. Yes, humans have been ramping up our production of renewable energy. The trouble is, we've also been ramping up our exploitation of fossil fuels. In other words, renewable energy has mostly just added to the energy mix (instead of replacing non-renewable energy).

And that brings me to the most rational road forward. Yes, we should invest in renewable energy — far more than we're doing today. But we shouldn't just wait for this technology to replace fossil fuels. Instead, the fossil-fuel business should be actively dismantled.

How?

The most severe option would be to make fossil fuels illegal. Although such a ban might one day happen, today it seems far-fetched. More likely, the fossil-fuel business will die by a thousand cuts. We already have policies devoted to this task. Carbon taxes, for example, are supposed to discourage fossil fuel consumption by making it more expensive.

[12] For the company that goes bankrupt, its 'degree' of bankruptcy is irrelevant. But for *creditors*, the degree of bankruptcy determines how much of their credit they can get back.

What a carbon tax does not necessarily do, though, is make the fossil-fuel business less profitable. There's probably a good reason for that. When we dig into corporate profits, we poke the hornet nest of power. At present, it seems that few governments have the stomach to fight the corporate swarm.

And that brings me to the fossil-fuel industry's self-made mess. One of the most pervasive norms in capitalism is that if you damage other people's property, you're liable for the cost of repair. Speaking of damage, the oil-and-gas business promised to cleanup the mess that it created. But so far, it has reneged on this promise — largely because it has the power to do so. The result is that today, the oil business has a massive cleanup liability. In the case of Alberta, it's between $40 to $80 billion. If the Alberta government called in this debt, it would likely bankrupt the (conventional) oil industry.

Of course, the Alberta government could achieve the same outcome simply by *taking* $80 billion from oil companies. However, the optics of doing so are terrible. Language tells us why. If the government enforces cleanup costs, it is being 'repaid' for an outstanding debt. But if the government takes the money without pretext, we call that 'stealing'.

These optics matter because human society is built on morals. Today, many environmentalists feel that we have a moral obligation to end the fossil business. But for the bulk of the population, this moral is a hard sell.

Ironically, the fossil-fuel industry has backed itself into a corner where it is vulnerable to a much more pervasive moral: the belief that you should cleanup your own mess. If we enforced this moral (which is, conveniently, also the law), the fossil-fuel business would likely be in deep trouble. It has dug its own grave and now stands on the edge. We just need to give it a push.

Appendix

Cleanup scenarios

Here are some possibilities for how Alberta's cleanup battle might play out. One option is that cleanup liabilities are never enforced. Sadly, this is by and large what's happened to date. Oil companies play accounting games to get their cleanup liabilities off the books. So far, the Alberta government (and the court system) have been willing to play along.

Another option is that the Federal government will finance the cleanup — a process that's already occurring. In 2020, the Federal government announced a $1.7 billion fund to cleanup abandoned and orphaned wells (Anderson, 2017). While the cleanup itself is good, this method of financing it is pernicious. The federal funding essentially tells oil companies that they can bolster their bottom line by ignoring their cleanup obligations. Then the federal government will come to the rescue and cleanup the mess. So the federal government is basically subsidizing oil companies' profits. That's bad.

Now let's get more radical. If the Alberta government called in the oil industry's cleanup liabilities, it would likely bankrupt the (conventional) oil business. If this enforcement had real teeth, we can imagine the government putting the whole oil patch into receivership. The government could then stop new drilling and drain existing wells. Whatever income that resulted would be put towards the oil-and-gas cleanup. If there was money leftover (and that's a big if), it could help fund a transition to renewable energy.

Acknowledgments

Thank-you to Regan Boychuk for getting me involved in this project, and to Eric Neilson for his work wrangling the oil-and-gas cleanup data, as well as his helpful explanations of the Enverus production dataset.

Thank-you also to the generous patrons who have supported this research: Hilliard Macbeth, Grace and Garry Fix, Pierre, Michael Rosenzweig, Steve Keen, Tim Ward, Norbert Hornstein, Tom Ross, Fernando, John Medcalf, Jeffrey Emmett, Jane k Willenbring, Mike and Joane Tench, Rob Rieben, and James Young.

Sources and methods

Data and code
The data and code used in this paper is available at the Open Science Framework: https://osf.io/rhnyz/

Oil-and-gas production

Oil-and-gas production data is from Enverus (formally known as DrillingInfo). For details about how I created well-level production curves from this data, see Fix (2022).

Cleanup cost data

Estimates for well cleanup costs come from an internal review conducted by the Alberta Energy Regulator. Following a freedom-of-information request, Regan Boychuk obtained the raw data from this review. This data consisted of private sector cleanup estimates for a variety of well scenarios. Boychuk then worked with Eric Neilson to apply the correct scenario to each well in Alberta. The resulting dataset estimates the cleanup costs of over 300,000 wells. For each well, I use the average of the low and high estimates.

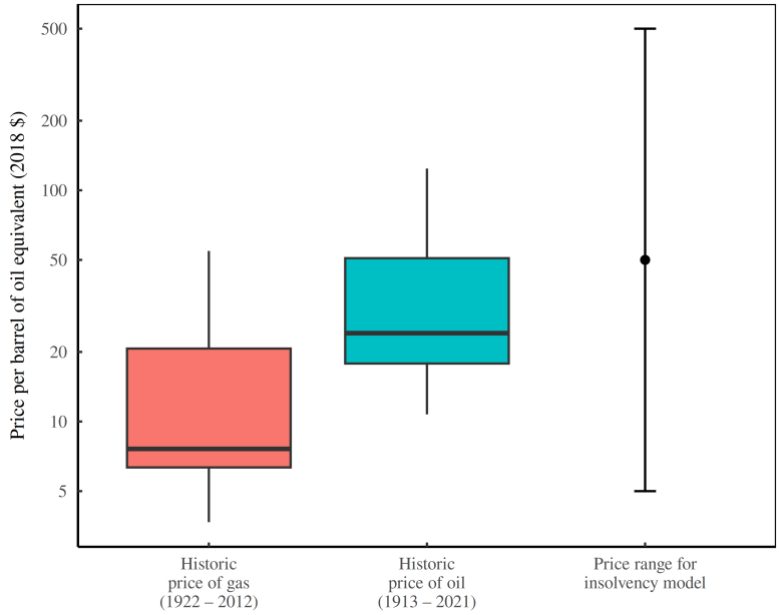
To construct the cumulative cost curve (Figures 8 and 11), I introduce each well's cleanup cost on the date the well was drilled. For simulated wells (those drilled after 2017), I assume a cleanup cost of $204,642 per well — the average of the historical estimates.

For more details about the cleanup estimates, see the Alberta Liabilities Disclosure Project's methods page: https://www.aldpcoalition.com/research

Historical oil-and-gas prices

Figure 13 gives you a sense for how my oil-and-gas price scenarios compare with historical data. Note that I've denominated prices in 'barrels of oil equivalent'.

Figure 13: Historical oil-and-gas prices. The box plots show the range of historical gas-and-oil prices, denominated in 2018 USD per barrel of oil equivalent. The error bars on the right show the price range used in my insolvency model. Oil prices comes from the BP Statistical Review 2021. Gas prices comes from Energy Information Agency, Table N9190US3a (wellhead price). I deflate both price series (to 2018 dollars) using the US Consumer Price index, obtained from the Bureau of Labor Statistics, series CUUR0000SA0.



References

AER. (2021). 2021/22 orphan fund levy - LLR and OWL programs. *Alberta Energy Regulator*. https://www.aer.ca/regulating-development/rules-and-directives/bulletins/bulletin-2021-09

AER. (2022). About the AER. *Alberta Energy Regulator*. https://www.aer.ca/providing-information/about-the-aer/who-we-are

Alberta. (2022). Alberta Heritage Savings Trust Fund: 2022–23 second quarter. *Alberta Government*. https://open.alberta.ca/dataset/ee01fe55-ac45-4bc9-b46e-34259ae6e3b9/resource/da420ac5-761d-4f7e-bde6-af38f10aea1a/download/tbf-2022-23-second-quarter-report-heritage-fund.pdf

Anderson, D. (2017). $1.7B to clean up orphaned and abandoned wells could create thousands of jobs. *CBC News*. https://www.cbc.ca/news/canada/calgary/federal-oil-and-gas-orphan-wells-program-1.5535943

Bellefontaine, M. (2018). $260B liability figure for abandoned energy infrastructure an "error in judgment": AER. *CBC News*. https://www.cbc.ca/news/canada/edmonton/alberta-energy-regulator-liability-figure-error-1.4888532

Boychuk, R. (2010). Misplaced generosity: Extraordinary profits in Alberta's oil and gas industry. *Parkland Institute*, 1–44. https://www.parklandinstitute.ca/misplaced_generosity

Boychuk, R. (2017). Who should pay to clean up the oil patch's Alberta mess? *Canada's National Observer*. https://www.nationalobserver.com/2017/03/31/analysis/who-should-pay-clean-oil-patchs-alberta-mess

Boychuk, R., Anielski, M., Snow, J. Jr., & Stelfox, B. (2021). The big cleanup: How enforcing the polluter pay principle can unlock Alberta's next great jobs boom. *Alberta Liabilities Disclosure Project*, 1–39. https://www.aldpcoalition.com/_files/ugd/6ca287_aebaacad30bd4

De Souza, M., McIntosh, E., & Bruser, D. (2018). Cleaning up Alberta's oilpatch could cost $260 billion, internal documents warn. *Global News*. https://globalnews.ca/news/4617664/cleaning-up-albertas-oilpatch-could-cost-260-billion-regulatory-documents-warn/

Finkel, A. (1989). *The social credit phenomenon in Alberta* (Vol. 12). University of Toronto Press.

Fix, B. (2020). As we exhaust our oil, it will get cheaper but less affordable. *Economics from the Top Down*. https://economicsfromthetopdown.com/2020/12/03/as-we-exhaust-our-oil-it-will-get-cheaper-but-less-affordable/

Fix, B. (2021). The ritual of capitalization. *Real-World Economics Review*, (97), 78–95.

Fix, B. (2022). A case study of fossil-fuel depletion. *Economics from the Top Down*. https://economicsfromthetopdown.com/2022/02/10/a-case-study-of-fossil-fuel-depletion/

Grandia, K. (2013). Alberta energy regulator Gerry Protti is the oil patch lobby's golden goose. *Huffpost*. https://www.huffpost.com/archive/ca/entry/alberta-energy-regulator-gerry-protti-is-the-oil-patch-lobbys-g_b_3233529

Johnson, T. (2017). Alberta shuts down Lexin Resources, leaving big mess to clean up. *CBC News*. https://www.cbc.ca/news/canada/calgary/lexin-resources-shut-down-1.3984287

Mallick, H. (2015). The Nixon of the North: How Stephen Harper ruined Canada. *Harper's Magazine*. https://harpers.org/archive/2015/10/stephen-harper-canada-nixon-of-the-north/

NBIM. (2023). About the fund. *Nbim.no*. https://www.nbim.no/

Paperny, M., Watson, J., & Frans, S. (2021). PricewaterhouseCoopers Inc v Perpetual Energy Inc 2021 ABCA 16. *Court of Appeal of Alberta*. https://www.canlii.org/en/ab/abca/doc/2021/2021abca16/2021abca16.pdf

Perez, S. (2020). Apple digs in heels over its App store commission structure with release of new study. *Techcrunch.com*. https://techcrunch.com/2020/07/22/apple-digs-in-heels-over-its-app-store-commission-structure-with-release-of-new-study/

Wilt, J. (2017). Norway's oil savings just hit $1 trillion. Alberta has $17 billion. What gives? *The Narwhal*. https://thenarwhal.ca/norway-s-oil-savings-just-hit-1-trillion-alberta-has-17-billion-what-s-gives/

Author contact: blairfix@fastmail.com

SUGGESTED CITATION:
Blair Fix, "How to Make the Oil Industry Go Bust", *real-world economics review*, issue no. 103, 31 March 2023, pp. 2-25, http://www.paecon.net/PAEReview/issue103/Fix

You may post and read comments on this paper at http://rwer.wordpress.com/comments-on-rwer-issue-no-103/

real-world economics review, issue no. 103
subscribe for free

Technological Change and Strategic Sabotage: A Capital as Power Analysis of the US Semiconductor Business

Christopher Mouré [York University, Toronto, Canada]

Abstract

Rapid technological change is often touted as a fundamental reality of capitalist societies. It is also presented as concrete evidence for the supposed progressive improvement of material well-being that characterises the capitalist system of social order. Since its emergence in the mid-20th century, semiconductor technology in many ways exemplifies this view. Yet the rapid advancement of semiconductor technology has also been accompanied by social conflict. The history of the technology is as much a story of frequent global chip 'shortages' and geopolitical disputes as it is one of exponentially growing computational power. The purpose of this study is to examine how the two sides of this story—progress and conflict—are linked. Starting from the theoretical political economic framework of capital as power, I put organized social power at the centre of this inquiry. I examine the behaviour of large semiconductor manufacturing firms in an attempt to uncover empirical relationships between capital investment, chip 'shortages', prices, and profits. Using quantitative and qualitative analysis, I find evidence that dominant semiconductor firms have engaged in systematic underinvestment in order to control chip prices for differential gain.

1. Introduction

In this paper, I examine the relationship between technical change and capitalist power in the US semiconductor business. The theoretical approach used is the capital as power framework, which argues that, in their quest for accumulation, capitalists seek to subjugate human creativity—including the creative and open-ended processes of technological change—to power. Furthermore, the capital as power framework argues that capitalists seek to *differentially* accumulate power—that is, relative to other capitalists—and the primary means of differential accumulation is 'strategic sabotage': the measured disruption of social, creative, and cooperative processes. One implication of this argument is that firms developing and selling new technology might accumulate by *limiting* technical change as well as *unleashing* or *propelling* it. If this is the case, and semiconductor firms must constantly attempt to subjugate technological change to the interests of business, does rapid technological change occur because of this fact or *in spite of it*?

Historically, the rapid pace of technological change in the semiconductor business was partly caused by *limits* placed on the ability of dominant electronics firms to sabotage the creative processes necessary for technical change. Specifically, a combination of military involvement and anti-trust laws preventing the monopolization of semiconductor technology through intellectual property unleashed 'industry' and encouraged fragmentation and rapid change over oligopoly and stagnation.[1] As a result of these restraints on business control, firms differentially accumulated primarily by coordinating the strategic limitation of the production volume of new chip technologies as they were introduced.[2] The lasting result of this strategy is that firms have been able to increase their differential profitability through temporarily increased prices, justified by the perception of a chip 'shortage'. Qualitatively, dominant semiconductor firms achieved effective control of semiconductor production capacity primarily through lobbying, business-government coalitions, and implicit cooperation between firms. However, the evidence suggests that while this strategy had some success, the speed of technological change, while propelled in part by firms' need to maintain competitiveness, was more of a *problem* for differential accumulation than an aid. This problem was particularly evident from the mid-1990s to the mid-2000s when the growth of new firms increased uncertainty regarding the future earning capacity of dominant firms. Finally, in the mid-2000s, dominant semiconductor firms responded to the centrifugal forces of technical change through a wave of mergers and acquisitions, resulting in a reduction in the volatility of their capitalization and a significant increase in their differential profitability.

2. Capital as power: a theoretical introduction

Capital as power diverges from neoclassical and Marxian political economy by arguing that profit is not a magnitude of utility or labour time but a manifestation of organized social power. As such, capital is not a *productive* entity but a symbolic representation of the power struggle among different capitalist groups as well as between these groups and the rest of the population, where the ultimate goal of these capitalist groups is the differential accumulation of social power (Nitzan & Bichler 2009, 218).

Building on the work of economist and social critic Thorstein Veblen, Nitzan and Bichler divide society into two distinct yet interdependent spheres: 'industry' and 'business'. According to Veblen, 'industry' consists of the collectively produced knowledge and creative activity of society. 'Business', on the other hand, is concerned only with profit. Crucially, 'business' uses private property law to restrain industry — an act of strategic exclusion that Veblen called 'sabotage' (Veblen 1908, 534–536). Following Veblen, Nitzan and Bichler argue that profit results not from the production of social goods but from a firm's ability to "strategically limit social creativity and well-being" (2009, 261).

Although business "does not and cannot make industry productive," it "can and does still 'propel' it" (Nitzan and Bichler 2009, 226). What does this mean? For Nitzan and Bichler, the logic of capitalism induces "human beings, organizations and institutions into a state of hyperactivity, constantly shaping and restructuring their interactions" (226). Insofar as that hyperactivity enlarges the scope of human wellbeing and "the inter-subjectively defined 'good life'" however, "it simply becomes a part of 'industry'"

[1] The meaning of the term 'industry' here is taken from the work of Thorstein Veblen, which is discussed below.

[2] In this paper I use the term 'chip', short for 'microchip', to refer generically to electronic components containing semiconductors. While the term "semiconductor" can also be used to designate any material that imperfectly conducts electricity (like silicon), I use it interchangeably with the term 'chip' to denote manufactured semiconductor components. Semiconductors in this context are manufactured products that use this conductive property to create an electronic on/off switch, called a transistor, which is the physical foundation of binary computing. Some semiconductors contain only a few transistors, while the most expensive ones, like microprocessors, can contain billions (Brown and Linden 2009, 7).

(226). On the other hand, such 'propulsion' is only profitable to the degree that it "interfere[s] with and partly hamper[s]" the creative and cooperative processes of 'industry'" (226). While business can unleash industry to a greater or lesser extent, it does so only in service of the future-oriented goal of maintaining or augmenting power over the creative processes of industry.

This hyperactive dynamic is central to understanding how capital accumulation can operate through sabotage and, simultaneously, through the rapid propulsion of technological change. For most of the history of capitalism, both rapid technological change and the strategic sabotage of technological change have existed together. It is possible that rapid technological change *in general* is actually one *effect* of the differential struggle for power over technology. In any case, capital as power theory suggests that contrary to the claims of semiconductor firms, the race for ever faster and more powerful microchips is not and cannot be the central goal of those firms, but only a means to the ends of business.

Because social power is always held and exercised in opposition to other groups, Nitzan and Bichler argue that it must be measured differentially, as a *relative* quantity, instead of in terms of maximization (2009, 18). What concerns capitalists, they propose, is the size of their power relative to other capitalist groups and relative to the underlying population. A common way to do this is to measure manifestations of power, particularly profit and stock price, relative to an average benchmark. The proliferation of financial indexes attests to the importance and commonness of this differential approach. Notably, differential accumulation can occur under conditions of stagnation as well as expansion. If a given firm shrinks 5 percent while the average firm shrinks by 15 percent, that firm has differentially *increased* its social power by 10 percent relative to the average firm.

The primary means of accumulation, Nitzan and Bichler propose, is to sabotage society to just the *right* degree – enough to earn differential profits, but not so much that earnings decline (2009, 236-237). This sabotage tends to generate resistance, and as a result, its strategic implementation often takes subtle, obscure, and incremental forms. Generally, the expectation of at least an 'average' return on investment is considered 'normal' in capitalist societies, and the associated social sabotage in its broadest sense becomes largely invisible (242). However, because sabotage is a socially negative phenomenon, it must often be *justified* in terms of some external, unavoidable necessity to minimize resistance.

Nitzan and Bichler argue that, in general, strategic sabotage takes two forms. First, business attempts to redirect industry toward more profitable ends. This may entail, for instance, investing in individually owned, fossil-fuel-dependent car transportation over electrified mass transit; or using proprietary pharmaceutical solutions to solve problems that have social or environmental causes that might otherwise be prevented (Nitzan and Bichler 2009, 234). These forms of sabotage are challenging to quantify and often turn on the assumption of counterfactual arguments (arguments about what might have happened if what actually happened had not). Thus, while they remain theoretically and practically important, they often make for a more challenging or indirect empirical inquiry.

The second form of sabotage Nitzan and Bichler outline is the systematic under-utilization of capacity. This limitation is easier to quantify and often yields surprising and counter-intuitive empirical results. This paper will primarily focus on this second form of limitation by examining whether semiconductor 'shortages' are possibly a result of this form of strategic sabotage.

Two other concepts are essential to our inquiry. Nitzan and Bichler argue that strategies for differential accumulation can be conceptually decomposed into two distinct categories: 'breadth' and 'depth' (2009, 328). A breadth strategy consists of expanding the size of the organization faster than the average, which can be achieved internally through 'green-field' investment or externally through mergers and

acquisitions.[3] Mergers and acquisitions are the more successful and preferred approaches because they expand a firm's *share* of production without increasing the overall level production in the sector as a whole (331). In contrast, green-field investment risks undermining differential profitability. If other firms competitively increase their output in response, tacit/open collusion between firms can give way to price wars (335).

In contrast to breadth, which focuses on the relative size of the organization, the strategy of depth consists of raising a firm's relative profit by increasing profit per employee faster than the average (or lowering it more slowly). While accumulation through depth is effective in the short term, Nitzan and Bichler argue that it is riskier in the long term. The problem is that because depth entails greater conflict, it is likely to meet stronger resistance from society and other capitalist groups (2009, 332). In addition, prolonged depth creates the opportunity for competitors to gain market share by setting lower profit margins. As a result, depth strategies are more often temporary fixes, and can result in unintended structural shifts in the landscape of power. This is especially true for the semiconductor business. 'Shortages' tend to occur periodically and for short periods of one or two years. Even so, this systematic depth strategy has brought new competitors into the fold and galvanized structural changes in the business over the long run.

3. Early dynamics of the US semiconductor business

In order to understand the contemporary state of the semiconductor business, it is important to understand its historical development. In particular, an analysis of the early years of the US semiconductor business offers insight into how the semiconductor business gained such an iron-clad reputation for dynamic technological change, and how this occurred in the context of the business logic of strategic sabotage. The following historical analysis details the role of the US military and government and the part of compulsory licensing of semiconductor patents in fostering a business sector oriented towards rapid technological advances – creating a relatively open and fragmented business landscape.

The role of the US military

The military began its involvement in the semiconductor business by creating it. During World War II, the military sponsored an extensive research program involving thirty to forty research labs to improve the silicon diodes used in radar (Flamm 1996, 29-30). After the transistor was invented in 1947 in Bell Telephone Laboratories (a participant in the government program), it became clear that the small size, greater reliability, and lower cost of transistors would lead them to replace the vacuum tube amplifiers that were widely used in electronics at the time (30). Military involvement in semiconductor research was directed through the Signal Corps Engineering Laboratory of the US Army (31). It started in 1950 and grew to "20 percent of total [semiconductor research] funding by 1952, and 50 percent of transistor work by 1953…stay[ing] at that level through 1955" (31). Flamm estimates that "about 25 percent of Bell Labs' semiconductor research budget over the period 1949-58 was funded by defence contracts, and all of the early production of Western Electric, the Bell System's manufacturing affiliate, went to military shipments" (31). The Signal Corps expanded its research funding beyond Bell Labs in 1955 and doubled the amount of its semiconductor funding to $1 million a year in 1956 (31). It also gave $50 million to engineering development between 1952-64, which focused on bringing technology beyond the prototype stage and into mass production (31). In 1956, "$15 million in contracts was appropriated,

[3] 'Green-field' investment refers to the internal expansion of production. For instance, this can include investing in new plant, property, and equipment, hiring employees, or expanding research and development efforts.

with funds flowing to virtually every semiconductor company in the United States" (31). The Signal Corps also funded the development of photolithography, which would become a critical process in the mass manufacture of integrated circuits (30). In short, the US military funded the research and paid for the resulting products at premium prices (24). Government influence was so widespread that a congressional report estimated that "if university and federal laboratory work were factored in, along with engineering development funds, and indirect R&D funding embedded in contracts to procure new devices at premium prices… the federal government paid for 85% of all US electronics R&D in 1959" (32-33). Because the military was interested in creating a large supply of high-performance transistors, many individual firms were paid to build factory capacity far in excess of orders (33).

The reliance on military and government funding only increased with the invention of the integrated circuit (IC) in the late 1950s.[4] While the companies that developed ICs rejected military funding for research and development to maintain control over the resulting intellectual property, almost all ICs in the early 1960s were used in defence systems and thus, the military, as virtually the only customer, remained the primary source of funding for IC development (34).[5] Tilton notes that defence production as a percentage of total production was 100 percent in 1962, 94 percent in 1963, and 85 percent in 1964 (Tilton 1971, 91). By 1968, the share was still 37 percent (91). In total, some estimates put federal funding at between 40 and 50 percent of all industrial semiconductor R&D from the late 1950s to the early 1970s (Flamm 1996, 36). While military funding for R&D dropped sharply off in the 1970s, the military remained a critical driving force in developing certain specialized areas of the technology (36). In addition, though the government share of semiconductor consumption declined, one source still estimated that "in the mid-1980s, purchases by defence agencies alone—direct and indirect—accounted for over one-quarter of US semiconductor shipments" (37).

Interestingly, the US military during the 1950s and 60s tended to favour new, untested firms (Flamm 1996, 32). For instance, "new firms (those with no background in the older vacuum tube business) accounted for 69 percent of military sales, and 63 percent of all semiconductor sales" (32). While Flamm argues this predilection was due to the military's quest for high-performance products, the logic may also have been to undermine differential power and keep the creative forces of 'industry' relatively unconstrained (32).

The pervasive influence and role of the US Military had three critical effects on the sector relevant to our inquiry here. First, the military's eagerness and deep pockets pushed firms to focus on high-performance rather than cost-effectiveness. This meant that the accumulation strategies of semiconductor firms were geared towards advancing the technology as quickly as possible in the hopes of landing a lucrative military contract. Second, military largesse in R&D funding lowered the risk of investment by subsidizing the costs of inventing the technology in the first place. Third, the preference for small firms may have undermined the differential power of established electronics firms, in a bid to 'propel' the industrial processes of technological change rather than restrain them. Almost none of the dominant vacuum tube firms became dominant in the US semiconductor business. At the same time, there is evidence that in Europe, the reliance on established electronics firms irreversibly slowed the introduction of semiconductor technology (Flamm 1996, 24-25).[6] In effect, the strategies for accumulation within the semiconductor business were heavily shaped by the US military's interests, making rapid technological change an unavoidable priority.

[4] Integrated circuits are semiconductors that contain several components, including transistors, on a single discrete chip.

[5] The first major application of ICs was in the Minuteman II guided missile (Flamm 1996, 34).

[6] See Malerba 1985 for an account of the historical development of the European semiconductor business.

The role of anti-trust and compulsory licensing

The second important factor in early semiconductor development in the US was a relatively open system of intellectual property (IP) controls. In David Noble's *America by Design* (1979), he argues that since the enactment of the Patent Act of 1836, the US patent system has vastly augmented the power of large corporations by granting temporary monopolies on new science-based knowledge (84-87). By the 1930s, he writes, it was clear that patents benefitted firms over the inventors themselves, whose control over their activities was arrested by "the compulsory signing away of patent rights of employees" (90). In addition, patent pooling between large firms effectively locked out new entrants (93). The result was a system that many saw as an impediment to the growth of technical and scientific knowledge and the diffusion of valuable technologies.

However, the 1940s brought a dramatic shift in the US Department of Justice's (DOJ) position on antitrust and "a new period of aggressive prosecution of corporate patent monopolies began" (Noble 1979, 88). This general context of strong antitrust enforcement is indispensable for understanding the emergence and development of the semiconductor business. Perhaps the most critical event symbolizing this era was the anti-trust suit brought against AT&T in 1949, culminating in a consent decree in 1956.[7] AT&T's liberal IP policies during and following the suit were instrumental in shaping the structure of semiconductor development (Grindley and Teece 1997, 12).

John Tilton offers a detailed account of the situation in *The International Diffusion of Technology: The Case of Semiconductors* (1971). "During the early fifties," he writes, "the patent and licensing policies of AT&T were the only ones of importance for a firm aspiring to enter the semiconductor industry…[because] AT&T was the pioneer, held the strategic patents, and possessed the vital know-how" (73). Royalties before the antitrust suit ranged from 0 to 5 percent (74). While Tilton argues that AT&T did not use its patent position to suppress competition, in the 1940s, the DOJ was pursuing a relatively aggressive antitrust program, specifically targeting firms holding large patent portfolios (Hart 2001, 928). The antitrust division was led by Thurman Arnold, who took over in 1938 and whose "stated objective was to convert antitrust… into a tool for 'breaking bottlenecks', including those that inhibited technological innovation" (928). After his appointment, Hart states that:

> Arnold's dramatic expansion of the use of consent decrees… allowed DOJ to establish the terms for settlement with defendants and excluded the judiciary from the process of resolving many cases. Over the course of the next decade, despite opposition in Congress and from big business and the military, Arnold and his followers moved antitrust policy in an increasingly deconcentrationist direction. Compulsory patent licensing, for instance, for the first time became a common element in antitrust settlements in the immediate post-World War II period. (928)

Under Arnold's supervision, "[the] DOJ soon filed suit against some of the nation's best-known high-technology companies, including Standard Oil of New Jersey, DuPont, General Electric, and Alcoa, and focused particularly on the patent holdings of some of these firms" (928).

The AT&T suit was opened in 1949 and ended in 1956 in a consent decree ordering Western Electric, which managed licensing for AT&T, to "license all existing patents royalty-free to any interested domestic firm… and all future patents at reasonable rates" (Tilton 1971, 76). In addition, Western Electric's semiconductor manufacturing operations were restricted to sales to the government and

[7] Both Bell laboratories and Western Electric were subsidiaries of AT&T. Bell Laboratories was the site of research and development, and Western Electric handled manufacturing and licensing.

AT&T sister companies, severely restricting the future growth of its semiconductor business (68). Even while the suit was ongoing, Tilton argues that it "must have influenced the company's policy of swiftly disseminating its new technology" (76). For instance, between the opening of the case and the consent decree, Western Electric lowered its maximum royalty rate from 5 to 2 percent and held several symposia in which its representatives exhibited and described the different technologies they had developed, including the transistor, diffusion, oxide masking, and other important developments (74-75).

Another crucial aspect of the consent decree was the stipulation that Western Electric could ask for a cross-licensing provision with its licensees (76). Because Bell Laboratories patents were so central to semiconductor technology, firms had a strong incentive to share their own patents with AT&T to gain access, and the proliferation of cross-licensing agreements only further sped up the process of information dissemination. In addition, due to the tough stance of the DOJ, most other semiconductor firms adopted a similarly liberal licensing approach out of fear of landing their own antitrust suit. In short, combined with the consent decree, the prevailing antitrust atmosphere had the effect of 'unleashing' the creative and cooperative forces of 'industry' by limiting the sabotage wrought by intellectual property monopolies. It contributed to a much freer flow of information and know-how, making the semiconductor business a more highly fragmented sector than it otherwise might have been.

A final and related factor is the high mobility of engineers, who often leave to start their own firms (Tilton 1972, 78). Bell Laboratories is particularly notable for defections, and "as with patents and licences, Bell Laboratories set a precedent for the industry's behaviour" (81). Because it could not compete in the commercial semiconductor market and could not offer the same lucrative contracts as other semiconductor firms, it took a more collaborative approach, maintaining "friendly informal relations" with former employees – compounding the integrative power of its cross-licensing agreements.

To recap: there were two significant reasons for the more collaborative and un-hindered development of semiconductor technology and its rapid advancement and dissemination in the early years of its development. First, the military's strong influence and funding shaped the efforts of the semiconductor business towards rapid improvements in quality and a focus on high-performance over cost-cutting. Even as military spending as a proportion of the semiconductor market shrank in the 1970s, it remained an essential support, both through procurement of high-powered computers and in targeted R&D. Second, the consent decree imposed on AT&T, as well as the broader atmosphere of strong antitrust enforcement against patent monopolies, unleashed industry by making the technological knowledge needed to create and design semiconductors a public resource (Grindley and Teece 1997, 13). The limitations on AT&T also led it to foster a cooperative, integrative approach to semiconductor research. Moreover, the atmosphere created by an aggressive antitrust division of the DOJ under Thurman Arnold meant that the consent decree set a precedent that other semiconductor firms would follow. Together, though not exclusively, these factors shaped the structure of the US-based semiconductor business as one focused on high-performance and rapid technological change, often regardless of the cost, and one relatively open to newcomers and to sharing the leading advances amongst one another. Furthermore, this analysis suggests that the 'unleashing' of industry was primarily shaped by factors outside and even openly antagonistic to the interests of the business. By restraining the sabotage of industry by business through the carrot (military contracts) and the stick (antitrust enforcement), different organs of the US government and military laid the foundation for a more fragmented business sector that *could not help* but propel technological change rapidly forward in its attempt to accumulate power. As the next section will attempt to show, this structure proved to be an enduring *problem* for differential accumulation.

4. Production, profit, and shortage: strategic sabotage in the semiconductor business

In this section, I present quantitative analysis showing the close relationship between chip production, capital expenditures, and differential profit. I argue that the price of semiconductors is closely and negatively related to the amount of productive capacity made available by semiconductor firms. This means that chip prices tend to increase when production capacity decreases – or, more specifically, chip prices fall more slowly when production grows more slowly. Second, it argues that the most powerful semiconductor firms, which I collectively label "Dominant Semiconductor Capital" have been able to differentially accumulate by restraining—within limits—the production of semiconductors and using the resulting atmosphere of 'shortage' to justify raising prices.[8] Inspired by a similar study by Nitzan and Bichler on the role of Middle East conflicts on oil profits, this section shows how 'shortages' tend to follow a period of differential *de*ccumulation when Dominant Semiconductor Capital trails the average rate of profitability.[9] These 'shortages' also tend to *be followed* by a "reversal of fortune," in which firms exceed the average (Nitzan and Bichler 2002, 236). In short, firms' differential earnings are largest when they successfully restrict production to the point of a perception of a 'shortage', which is used as a justification for an 'depth' accumulation strategy of differential inflation.

Power, production, and pricing

How can it be empirically shown that this thesis is correct, and that Dominant Semiconductor Capital can differentially profit by limiting chip production? One starting point for gathering data is the US Bureau of Labor Statistics (BLS), which publishes detailed annual production data. Figure 1 uses BLS data to compare the rate of change in US semiconductor production volume and the rate of change of the sector's price deflator. The price deflator inflation is a proxy for overall price change. It estimates how much of the change in the dollar value of production is a result of 'pure' price changes, as opposed to a change in 'real' output.

The BLS data must be used with caution, as standard measures of 'real' output generally do not correspond to an actual quantity of goods but only to the monetary value of those goods, adjusted to remove so-called 'pure' price changes. Moreover, there are certain theoretical and methodological problems with determining when a price change is 'pure' (i.e., a result of 'supply' and 'demand' factors) and when it is the result of a difference in the quality of the product.[10] For instance, ICs have relatively short product life cycles, with new chips (that can have different physical characteristics and be intended for different uses) coming on and off the market in as little as 18 months. As Nitzan notes, "whenever the nature of the commodity changes, the measurement of such changes in 'quality' is crucial for price and quantity calculations" (1992, 156). This makes reducing the aggregate output volume of the entire manufacturing sector to a single monetary value difficult, if not impossible. In addition, the methods of the BLS in dealing with the price/quality question are largely opaque and are approached exclusively from a neoclassical economic perspective. This presents a problem for researchers critical of the fundamental assumptions of neoclassical theory because "the predisposition of price and quantity data

[8] For a detailed explanation for how I measure Dominant Semiconductor Capital empirically, see the Appendix at the end of the paper.

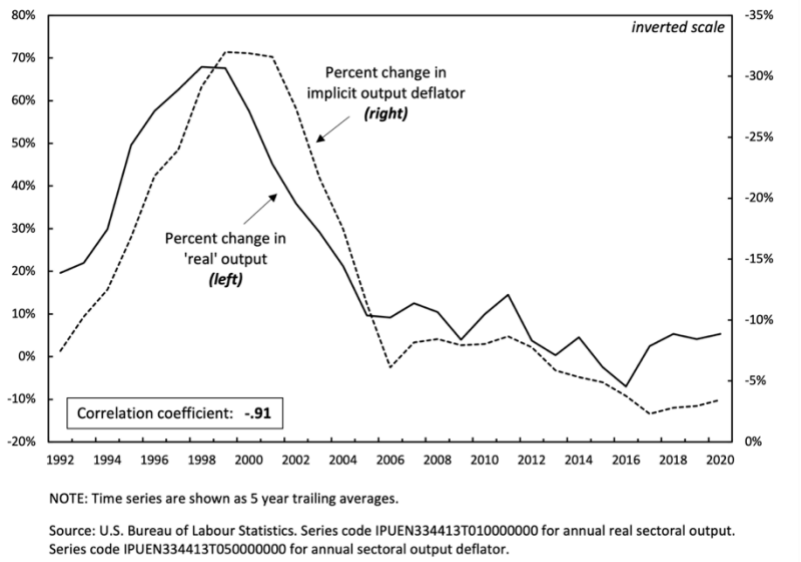
[9] In Nitzan and Bichler's (2002) work, they show how Middle East conflicts have played a similar role to semiconductor 'shortages', by causing the perception that oil production is being or will be disrupted and, in turn, justifying the raising of oil prices. Unlike semiconductor 'shortages', however, the results of this strategy have been immeasurably more tragic.

[10] See Nitzan 1992, Ch. 5 for a discussion of the price/quality problem.

toward the neoclassical economic outlook means that these data may not be altogether suitable to test the neoclassical outlook against competing frameworks" (158). For example, if 'pure' price changes are imputed to 'supply' and 'demand' factors, and this explanation is considered to lack factual information, what does this mean for the validity of the measure of 'pure' price change? Nonetheless, even after heavily discounting the meaningfulness of the data, it is interesting that they still appear to confirm the thesis that production and price changes are inversely correlated.

The very tight negative correlation between change in output volume and change in price shown in Figure 1 (-.91, note the inverted right scale) suggests that as the rate of *increase* in production slows, so does the rate of *downward* shift in price. This correlation implies that semiconductor price changes have an inverse relation to changes in production volume.

Figure 1. Percent change in 'real' output and percent change in prices, U.S. semiconductor manufacturing



NOTE: Time series are shown as 5 year trailing averages.

Source: U.S. Bureau of Labour Statistics. Series code IPUEN334413T010000000 for annual real sectoral output. Series code IPUEN334413T050000000 for annual sectoral output deflator.

If the growth rates of prices and production have an inverse relationship, what about the relationship between differential profit and investment in new production? Figure 2 compares the rate of change of differential profitability to the rate of change in capital expenditures for Dominant Semiconductor Capital (again, note the inverted right scale). Differential profitability is measured here as the percent deviation of Dominant Semiconductor Capital's return on equity (ROE) from the average ROE of the Compustat 500.[11] (From here on, and unless indicated otherwise, the measures for both Dominant Semiconductor Capital and the Compustat 500 are computed as overall weighted group averages.)

[11] The latter measure is calculated as the average return on equity of the largest 500 firms by market capitalization in the Compustat Capital IQ North America database. Return on equity is calculated as net income divided by total common equity.

Figure 2. Percent change in differential return on equity and percent change in capital expenditures, Dominant Semiconductor Capital

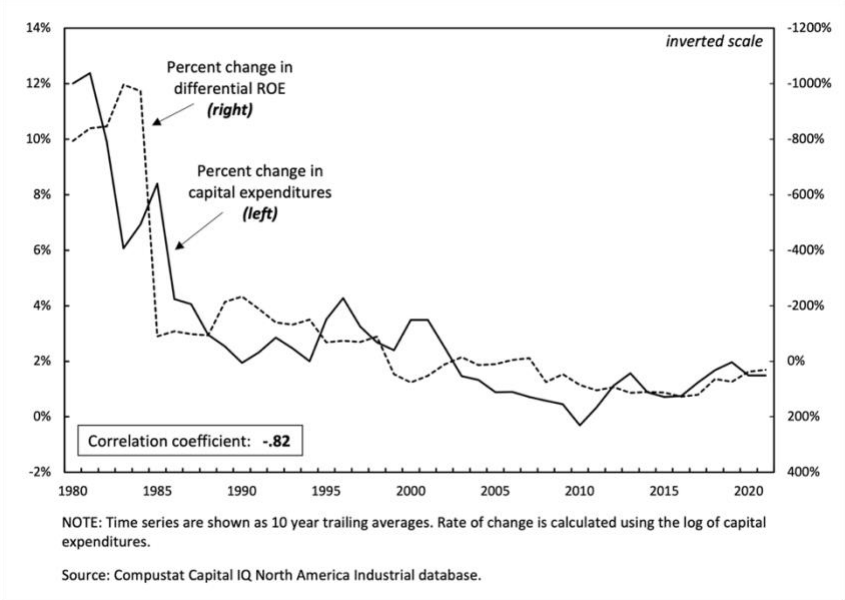


Figure 2 shows a significant negative correlation between the rate of change in differential ROE and the rate of change in capital expenditures (-0.82). The fact that it is negative is counter-intuitive from a neoclassical perspective. If profit is a 'cost' of production, logically it should increase faster with a more rapid increase in production, and vice versa when production decelerates. In addition, both profit and production should increase when there is an increase in 'demand'. According to neoclassical theory, this means that in a perfectly competitive market the growth of profit and production are likely to move together. However, whereas neoclassical economics focuses on absolute profit growth, in a landscape of shifting prices and antagonistic business relations, what matters is not absolute but *relative* return on investment. What Figure 2 shows is that there is a strong correlation between differential returns and *restraint* on new investment. From a capital as power perspective, this evidence is not as surprising. Capital as power theory argues that capitalists seek not absolute but differential returns, and that, consequently, they do not primarily seek to expand production but rather to subjugate output to their own differential goals. By strategically limiting production (i.e., by creating more 'scarcity'), capitalists can charge higher prices for their products. Conversely, if production expands too quickly, they are liable to lose control of pricing, resulting in lower relative prices and lower relative profits, and often lower differential returns. The need to strategically limit production explains the negative correlations shown in Figure 1 and Figure 2.

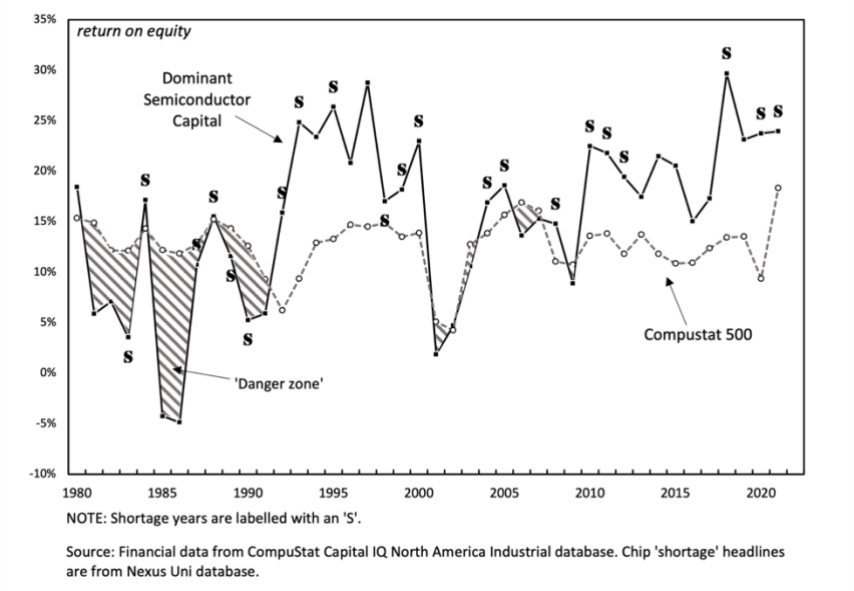
Differential profit and the perception of shortage

The next stage of the empirical investigation is to look at the relationship between production and profitability in the context of 'shortages'. If strategically reducing investments in new productive capacity can result in higher prices and higher differential rates of profits for semiconductor firms, it is plausible that 'shortages' are not a product of a supply-demand imbalance, as neoclassical theory would have it, but a predictable result of Dominant Semiconductor Capital's struggle to differentially accumulate.

The following analysis, as well as Figures 3 and 4, are inspired by Nitzan and Bichler's work studying the accumulation of arms and oil producers in relation to Middle East wars (see Nitzan and Bichler 2002, Ch. 5). Figure 3 compares the ROE for Dominant Semiconductor Capital and the ROE for the

Compustat 500, with labels marking 'shortage' years.[12] The shaded areas denote 'danger zones' when Dominant Semiconductor Capital experienced differential *de*cumulation. Figure 4 shows the percent deviation of Dominant Semiconductor Capital's ROE from the Compustat 500 average, again, with marked years of 'shortage'.

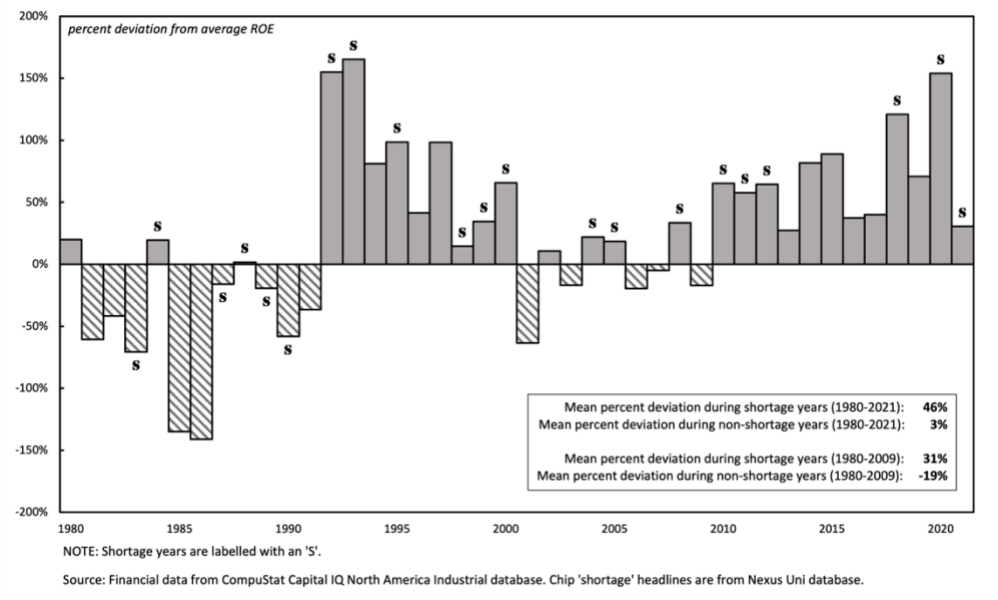
Figure 3. Return on equity and semiconductor shortages, Dominant Semiconductor Capital and Compustat 500 average



The pattern in Figure 3 is similar to the one found in Nitzan and Bichler (2002, Figures 5.7 and 5.8, 236-237). If we define a 'danger zone' as an uninterrupted period of a year or more in which Dominant Semiconductor Capital firms trail the average ROE, all but one danger zone between 1980-2021 ended with a perception of a 'shortage' the following year. In other words, the change in fortune from trailing to beating the average is almost always accompanied by an atmosphere of shortage. In addition, most years of differential accumulation are accompanied by 'shortage' (61%), in contrast to years of differential decumulation, in which 'shortages' are less common (28%).

[12] For a detailed discussion of how I measured 'shortages' or rather the perception of a 'shortage', see the Appendix at the end of the paper.

Figure 4. Differential profitability and semiconductor 'shortages'



In addition, as shown in Figure 4, there are substantial differences in differential profitability between shortage and non-shortage years. From 1980 to 2021, Dominant Semiconductor Capital exceeded the average by 46% during shortage years, while during non-shortage years, Dominant Semiconductor Capital only barely met the average. Between 1980 and 2009, Dominant Semiconductor Capital exceeded the average by 31% during 'shortage' years and trailed the average by -19% during non-shortage years. The 1980-2009 period is highlighted because, after 2009, there is evidence that the differential profitability of Dominant Semiconductor Capital was increasingly the result of a different strategy, that of mergers and acquisitions. This transition is examined below, in the section "Centrifugal forces."

The evidence suggests a close relationship exists between the strategic limitation of chip production, differential profitability, and perceptions of a chip 'shortage'. On the one hand, it could be that a shortage is an *unintended* effect of 'over-shooting' the collective reductions in chip production undertaken by Dominant Semiconductor Capital. On the other hand, the periodic regularity of shortages and their correlation with increases in profitability suggest an alternative: that the creation of the *perception* of shortages plays a crucial role in justifying the rise in chip prices from which firms differentially profit. As noted above, differential prices tend to generate resistance and reaction because they *redistribute* income. Framing the conscious reduction in investment and its consequent 'shortages' as an issue of supply and demand (i.e., outside any one firm's control) obscures the role of power in this process.

5. The 1986 US-Japan Semiconductor Trade Agreement (STA)

Profiting from differential inflation requires a minimum of cooperation, because one firm or group of firms can easily undermine the others by lowering prices and expanding market share. One of the most significant and well-known historical examples of such cooperation occurred as a result of the US-Japanese chip war of the 1980s. Recounting this history illustrates qualitatively and concretely one way the process of strategic sabotage in the US semiconductor business has unfolded.

While the popular narrative is one of US chip firms valiantly defending US chip production from government-backed Japanese firms, it was the US government that often "appeared to have been the

cause and not just the effect of changes in the competitive conduct of Japanese semiconductor producers" (Flamm 1996, 127). Although the level of strategic cooperation of firms and governments to consolidate control over the production of chips reached a high-water mark in 1986, calls for government assistance on behalf of the US semiconductor business began as early as the 1950s (127). As early as 1959, the Electronics Industries Association (EIA) "petitioned the Office of Civil and Defense Mobilization…to impose quotas on Japanese transistors" (128). The EIA at that time was split between electronics components producers and consumer electronics producers (129). On one side, components producers like semiconductor firms "favoured protection against consumer electronics imports from Japan—in which the bulk of inexpensive Japanese components entering the US market was embedded" (129). On the other, consumer electronics producers were worried about market access and were generally more deeply embedded in the Japanese market itself: some marketed Japanese products in the US for Japanese partners (Motorola and General Electric), some collected royalties from Japanese licensees (RCA), and others had significant ownership stakes in Japanese companies or controlled domestic Japanese subsidiaries (129). Thus, in 1968, the EIA ended up "testifying on both sides of trade policy issues" like protective tariffs (129).

By the 1980s, trade frictions between the US and Japan were compounded by the fact that, after embarking on a large-scale project of technological catch-up, Japanese semiconductor producers had surpassed US chipmakers in certain critical areas of production (Flamm 1996, 100). Central to this effort was the VLSI project, a series of multi-firm R&D subsidies organized by Japan's Ministry for International Trade and Industry (MITI) that "accounted for almost 40 percent… of Japan's national IC R&D effort in the late 1970s" (97-98). Out of the 22 significant results of the VLSI project, the majority were improvements in process technology that reduced costs and improved the quality of existing technologies (100). Consequently, Japanese chip producers vastly improved chip yield (the number of working chips produced in a single silicon wafer), allowing them to make chips at a much lower cost than their American competitors, particularly in DRAM, or dynamic random-access memory, manufacturing.

The creation of the Semiconductor Industry Association (SIA)—the new "lobbying arm" of large US semiconductor firms—in 1977 was a direct consequence of the growing awareness by US chip firms of the significant gains Japanese producers were making in international markets (Flamm 1996, 138). As Irwin notes, the SIA reflected the common interests between otherwise openly antagonistic firms:

> Each of these firms competed fiercely with one another on certain dimensions—suing each other over alleged patent violations, for example, or even conducting espionage against one another—but they could agree on several common policy objectives, such as obtaining greater patent protection for chip designs, improving the tax treatment of R&D investment, and heightening political awareness of the emerging Japanese competition. (Irwin 1996b, 21)

The organization successfully convinced the US Senate to "order the US International Trade Commission (ITC) to launch an informational investigation into the competitive position of the US semiconductor industry" by arguing that the Japanese were poised to steal market share from US producers through illegal trade protections (139). While evidence of illegality was thin, generous political donations made by the four largest chip firms during this period no doubt added gravity to the urgency of their claims (Irwin 1996b, 21). Irwin notes that "disbursements of these PACs [political action committees] appear to be related to the trade dispute with Japan: payments totalled $354,318 at the peak of the dispute in 1985-86, 40 percent higher than in 1983-84 and 17 percent higher than in 1987-88 after trade tensions had simmered down" (1996b, 22).

By the end of 1979, during a 16K DRAM 'shortage' in the US, Japan's three largest chip producers had managed to capture 40 percent of the US chip market (Flamm 1996, 139-140). The SIA argued that the Japanese had achieved this capture by 'dumping' using a two-tiered pricing structure: keeping prices high in the Japanese domestic market and exporting chips 'below cost' or 'below fair value' to gain market share in foreign markets (141). The evidence for this accusation, Flamm notes, was scarce and primarily anecdotal (138). Despite the lack of evidence, US chip producers soon began to complain that "the Japanese were selling below the cost of production in *both* [US and Japanese] markets" (Flamm 1996, 142).[13] Fearing anti-dumping actions by the US Trade Representative (USTR), Japanese producers suspended sales to the US spot market for DRAMs and announced plans to open US production facilities "in a complementary bid to reduce trade frictions" (143).[14]

Pressuring Japan

Political pressure on Japanese producers to reduce exports to the US and raise prices increased steadily through the early 1980s. In 1981, Japanese producers introduced new 64K DRAMS ahead of U.S producers and "by early 1982…the Japanese share of the US 64K DRAM market [stood] at about 70 percent" (Flamm 1996, 148-149). Again, US producers pressed the Commerce Department "to investigate charges that the Japanese were selling 64K DRAMS below 'fair value'" (149). In response, Washington warned the Japanese government that the "Commerce Department might begin to 'monitor' Japanese import prices" (149). However, as the investigation was set to begin, "DRAM prices suddenly doubled, Japanese suppliers began rationing US customers, and it was reported that Japanese companies were cutting back US exports in order to blunt moves toward trade restrictions on DRAM imports" (149). Companies confirmed to reporters that "they were reducing US exports to alleviate trade friction" at the behest of the Japanese (and US) governments (149-150). In 1983, another 'shortage' appeared within a year after introducing export reductions, and chip prices shot up (151). Ironically, following the 1983 price increase, US producers immediately began accusing the Japanese producers of collusion and price-fixing, and the US Justice Department opened an antitrust investigation into 'excessively high prices' (152). While the case eventually faded away, the mixed and contradictory accusations reportedly left Japanese producers somewhat confused (151). When the 1983-1984 'shortage' subsided, criticism of the Japanese returned with force. Antidumping petitions were filed against three types of chips, while a private antitrust suit and several complaints of unfair trade practices (under the infamous Section 301 of the 1974 trade act) were filed in 1985 (160). All but the antitrust case was suspended with the negotiation and signing of the Semiconductor Trade Agreement (STA) in September of 1986 (160).

The 'shortage' of 1987-1990

The 1986 STA established export price controls for several kinds of semiconductors and called for extensive monitoring of DRAM and EPROM prices by the Japanese government and a host of other products. The agreement stipulated that the Japanese government would "take appropriate actions

[13] One of the most curious charges was 'quality dumping' (Flamm 1996,145). Japanese firms had successfully improved their chip manufacturing standards to the point that US chip customers reported defect rates of Japanese DRAMs at "one-half to one-third those experienced with comparable American products" (145). Instead of denying the truth of these reports, US chip producers "charged that sales of higher quality Japanese products at the same price as American products reflected a form of 'dumping'" (145).

[14] A spot market is a more informal market for chips based on short-term contracts and populated by trading firms, manufacturing firms, and customers.

available under law and regulations in Japan, including ETC [the Export Trade Control ordinance], in order to prevent dumping," and implementation moved quickly (Flamm 1996, 177). MITI suspended new export licences for 256K DRAMs and advised Japanese producers to set export prices at or above 'fair value', which was dictated by the US Commerce Department (177). It also set up a Forecast Committee, "whose task it was not only to publish forecasts but also to help in 'correction of imbalances' between Japanese supply and demand for products covered by the price monitoring framework" (177-178).

At first, producers found "creative mechanisms to evade the newly minted MITI controls," and many firms did not cut back production to the levels encouraged by the Forecast Committee (Flamm 1996, 179). In particular, TI Japan and NEC, the largest producers of 256K DRAMs, were "reluctant to follow MITI's new 'guidance' on production and export volumes" (180). However, it is possible that a turning point occurred in March 1987, when President Ronald Reagan declared that "prohibitive (100 percent) tariffs would be imposed on $300 million worth of imports from Japan… [the] largest and the first [unilateral retaliatory trade] action against this US ally in the postwar period" (Irwin 1996b, 11). As Irwin notes, "the sanctions were crafted to hit the exports of the principal Japanese semiconductor producers—such as NEC, Toshiba, Hitachi, and Matsushita—but not entail significant consumer losses" (53). More than anything, these measures may have compelled Japanese chip producers to follow the 'guidelines' set by MITI.

Eventually, producers began to comply with the pressure. For instance, TI Japan announced it would "slash its output of 256K DRAMs by 13 percent to comply with MITI's wishes," and NEC announced it would cut production by 40 percent (183-4). By the spring of 1987, "both the production and export of DRAMs by Japanese companies had been placed under fairly tight MITI controls," and prices began to rise worldwide (185). As Flamm notes, "annual rates of change in price hit all-time historical highs in 1987 and 1988, after the STA went into effect," and in a business defined by constant falling prices, *positive* changes were recorded for the first time for memory chips in 1988 (237). Flamm estimates that the "'guidance' supplied to producers in restraining investment levels continued at least through early 1988, and therefore probably affected supply through at least 1989" (272).

By early 1988, rising prices brought complaints by chip users, as well as fears of another 'shortage' – one that was beginning to be perceived in both the US and Europe (Flamm 1996, 192). Even as prices continued to climb, however, "MITI's control framework was extended into new areas" (192). For its part, "the increasing signs of shortage satisfied the American government that MITI had acted forcefully to increase chip prices in worldwide markets" (193). However, to counter public pressure from the chip users, who were complaining about the high prices, the US again "switched its public posture to one of encouraging MITI not to restrict chip production" (193). In truth, however, the government was "less than unequivocal…officials were not eager to see Japanese firms increasing their chip capacity to meet the looming shortage, preferring to see American companies 're-enter' the DRAM market" (194). Far from acting to remove limits, "American trade negotiators continued to press MITI to limit investments by Japanese firms in new capacity well into 1988" (195).

Due to increasing complaints by Japanese chip users, "MITI became considerably more reluctant to spell out the precise nature of its actions in public" and began to advocate for firms to "extricate themselves from the inclination toward excessive competition" rather than relying on the guidance of MITI (Flamm 1996, 184-185). The idea that it was "indispensable for manufacturers to make their own efforts…to establish prices in accordance with the balance of supply and demand" was increasingly promoted both by private firms and the government (186). For instance, one 1987 MITI report "called for a considerable amount of coordination among rival firms in the semiconductor industry" and "specifically called on semiconductor producers to cooperate in planning investments and in matching

production to forecast levels of supply and demand" (187). In effect, both government and businesses agreed to 'privatize' the price and production controls implemented through the framework of the STA.

The uncanny accuracy of production 'forecasts'

Perhaps the most obvious evidence that MITI's production forecasts acted as semi-compulsory guidelines was their remarkable accuracy during the 1987-1989 'shortage' years (Flamm 1996, 201). For instance, between the second and fourth quarter of 1987, "forecasts issued for 256K, and 1M DRAM production levels three and six months out typically fell within 10 percent of actual output" (197). This level of accuracy continued even after "the invisible hand, rather than the government's, ostensibly ruled the market," attesting to the success of the privatization of price controls (197). Some analysts argued that the shortages were due to other factors than the purposeful restriction of production toward the aim of price increases (197). One prominent argument was that "unexpected yield problems were a major factor in the shortage then developing [in 1988]" (197). However, Flamm points out that "if such yield problems, in the aggregate, played a significant role in creating shortages…it is hard to see how they could have been unexpected, given the accuracy of the MITI forecasts" (200). "A continuous history of coming within 10 percent of three-and six-month production forecasts," he reasons, "in a product requiring over two months of processing on the production line, suggests that unanticipated yield problems could not have been a major issue" (200).

The entrance of Samsung and the second STA

The 'shortage' also produced centrifugal forces: principally, it allowed emerging South Korean chip producer Samsung, who was "not subject to the political and legal pressures faced by Japanese chip-makers," to undercut DRAM prices and vastly expand their market share. By the time the shortage atmosphere had dissipated, Samsung had become "the largest producer of 1M DRAMs in the world" (Flamm 1996, 222). On the other hand, Samsung's successful expansion incurred the ire of its competitors, and in 1990 Micron Technologies (one of two remaining US DRAM producers) "raised the possibility of a dumping suit against Korean vendors" (226). Like the Japanese, Samsung quickly obliged and cut DRAM production by 20% (226). In addition, "dumping cases against Korean exports of DRAMs were filed in the European Community (EC) in 1991 and in the US in 1992" (224). "Faced with stiff antidumping duties," writes Irwin, "the Korean industry and government proposed in January 1993 a bilateral semiconductor trade agreement fashioned on the earlier one with Japan," in which "the Korean industry promised to monitor prices of export sales to the United States" and ensure "demonstrable and measurable results in terms of increasing sales in Korea of US semiconductors and semiconductor equipment" (Irwin 1996b, 60).

In June 1991, a second STA was signed, replacing the one set to expire the following month (Flamm 1996, 223). This agreement removed the 'fair market value' pricing floors, introduced a fast-track antidumping procedure, and retained the extensive price monitoring system. Flamm argues that practically speaking, little changed (224). MITI continued to collect industry data and publish 'forecasts,' and "thus a variety of well-established mechanisms designed to constrain pricing in world semiconductor markets continued in their original or revised form" (224).

What can be concluded from these events? First, the events of the 1980s show that the notion of 'shortages' is unexplainable without reference to the broader *power* dynamics of the sector. Competition from Japanese producers was countered by coordinated political pressure to reduce competition and restrain production (Flamm 1996, 206). While coerced initially, Japanese firms quickly understood that

they had as much to gain from higher margins as US producers. As one Japanese executive put it: "since the Semiconductor agreement, we [Japanese DRAM makers] have moved from competing for market share to market sharing" (quoted in Flamm 1996, 215). The profits from differential inflation could only have been achieved through a coordinated effort to curtail production.

Second, the emergence of the SIA and the success of its lobbying efforts marks the emergence of a loose 'distributional coalition' (to use Mancur Olson's term) of semiconductor firms and government organs that coordinated the limitation of production to increase chip prices. If the correlation between 'shortages' and the differential profitability of Dominant Semiconductor Capital is any indication, the development of this coalition was far more critical than any initial profits gained from the DRAM 'shortage' of 1987-1990. Irwin notes, "only two US merchant firms (TI and Micron) remained in the DRAM market to benefit from the antidumping actions. DRAM sales reportedly accounted for as much as 60 percent of TI's profits in 1988, and Micron's sales rose by a factor of six between 1986 and 1988" (1996b, 11-12). On the whole, the creation of this ongoing government-business coalition was much more valuable than any initial monetary gains of these two firms. As Irwin argues, "such a sectoral agreement [as the STA] is attractive from the perspective of virtually any import-competing industry because it virtually guarantees the institutionalization of trade policy for that industry" (11). Moreover, "once the agreement was in place, it required monitoring and at some point, renewal, or renegotiation… [providing] a natural rationale for ongoing contacts between the industry and the government, providing the industry with easy access to key policymakers and allowing close industry-government ties to develop" (12).

Third, compounding the cooperation of governments and sanctioned by them is the open exchange of price and production information, for example, through the 'demand forecasts' published periodically by MITI and by organizations like the SIA. Publishing these forecasts allows producers to adjust new capacity investments to avoid unprofitable 'overproduction' without officially breaking the law. In addition, these structures of dominant power tend to recede into the background of the public consciousness as the political contention engendered initially by the creation of such a coalition is replaced by the matter-of-fact designation of the semiconductor business as a 'strategically important' industry.[15]

6. Centrifugal forces

In their 2002 book, *The Global Political Economy of Israel*, Nitzan and Bichler track the global shift from depth to breadth starting in the late 1980s (294). This transition, they argue, rested on three "breadth-related poles": *capital decontrols*—referring to the increasing opening up of national markets to foreign investment and ownership, particularly under free trade agreements like NAFTA and the EU—*privatization,* and expansion into '*emerging markets'* (295). While each of these certainly affected the shift to breadth within the semiconductor business, here I focus on a fourth set of factors—the centrifugal forces within the semiconductor manufacturing business—and their relation to Dominant Semiconductor Capital's shift to a breadth strategy in the mid-1990s. I argue that centrifugal forces at work in the US semiconductor business led to a steady acceleration in the growth of new firms between approximately 1985-2005. This growth first destabilized the differential power of dominant semiconductor firms, leading to increased uncertainty about their relative power. Subsequently, the rapid growth of new firms became the basis for a new breadth period, as large firms accelerated the

[15] For a discussion of whether or not the semiconductor business is 'strategic', see Flamm 1996, Ch. 7.

pace of corporate amalgamation, stabilizing their relative power and vastly increasing their differential earnings in the process.

As mentioned above, mergers and acquisitions are reliable forms of differential accumulation. As Nitzan and Bichler note, merger and acquisition activity "kill[s] three birds with one stone: it directly increases differential breadth [organization size]; it indirectly helps to protect and possibly boost differential depth (relative pricing power); and it reduces differential risk" (2009, 330). In addition, the *integrative* nature of information technology—reliant upon interoperability, the growth of networks, and its capacity to link social activity across geographic space instantaneously—lends itself to mergers and acquisitions (Nitzan and Bichler 2002, 294). Due to this integrative tendency, the expansion of the 'network' constantly introduces "new players, new forces, and new rules," which "if left unattended, tend to destabilize established power and undermine profit" (294). Nitzan and Bichler argue that "the most common way of containing these centrifugal forces is through the centripetal, counter-force of corporate amalgamation" (294-295). Thus, a breadth strategy will be more critical in a rapidly changing and expanding business landscape than in one which is more predictable, growing slower, and less dynamic. If this is the case, one would expect the semiconductor business to be an early adopter of a breadth strategy of mergers and acquisitions. However, it was not until the mid-1990s that acquisition spending rose to close to the corporate average (see Figure 5 and Figure 8). Why did it take so long for the semiconductor business to adopt such a strategy?

One possible reason is that there were simply too few firms to take over. Before 1982, less than 20 semiconductor manufacturing firms were listed in the Compustat North America database. High start-up costs and a small, mostly government market also meant that mergers and acquisitions, when they did happen, were minor in relation to other expenditures. In addition, non-US-based chip producers tended to be large, vertically integrated consumer electronic firms and government-backed 'national champions' – meaning they were too large, legally protected or both. These firms were neither available nor vulnerable to acquisition until the 2000s, when national economies 'globalized', capital controls were deregulated, and many large conglomerates 'spun off' their semiconductor operations into separate businesses (McClean 2011, 2-3).

However, despite the initial high concentration of firms, as computing grew in importance to the industry as a whole, and as semiconductor firms continued to experience the appearance of periodic shortages, important centrifugal forces generated more and more new firms. For one, the frequent appearance of shortages, along with higher prices, presented opportunities for new competitors to enter and gain market share by undercutting the other large firms (e.g., the case of Samsung).

Second, high prices and concentrated power are strong incentives for institutional chip *customers* to encourage greater fragmentation in the sector, and enterprising computing firms like Apple often actively seek to undermine the power of large chip producers like Intel. For instance, in 1987, Apple worked with firms VLSI and Acorn to create a jointly owned, independent company to design the chip for its Newton personal digital assistant (Nenni and McLellan, 2013, 35). While the Newton was not a commercially successful product, the joint venture, ARM, became one of the largest suppliers of semiconductor IP and the standard microprocessor architecture in mobile phones (170-171).

Third, the speeds at which semiconductor technology, and information technology in general, were developing, presented new opportunities for leading-edge or highly specialized producers to capture market share as new industrial uses emerged. For instance, as computing power increased and became cheaper and smaller, the range of services to which computation could be put grew. Despite these opportunities, large firms like Texas Instruments and Intel, which in the 1980s controlled every aspect of the chip-making process from design and manufacturing to packaging and testing, tended to

focus on mass production of only the most profitable products. The growing gap between these firms' product lines and the diversity of industrial needs opened up opportunities for smaller companies to specialize in these riskier and less profitable markets. A result of this increasing diversity of custom chips was the emergence of 'fabless' semiconductor firms, who design but do not manufacture their chips. The growth of fabless firms, which initially operated by renting excess capacity from existing manufacturers, only accelerated with the introduction of contract manufacturers called 'foundries'. Semiconductor foundries do not design their chips but devote their operations to producing chips designed by others. The most important of these is the Taiwan Semiconductor Manufacturing Company (TSMC), founded as a joint venture between the Taiwanese government and Dutch electronics manufacturer Philips—another example of a chip user acting as a centrifugal force—and other private investors (Nenni and McLellan 2013, 76). TSMC has become one of the semiconductor business's largest and most powerful firms (83).

In short, many strong centrifugal forces threaten to destabilize the differential power of dominant semiconductor firms.[16] The following analysis argues that the acceleration of these forces in the 1990s destabilized the power of Dominant Semiconductor Capital, giving rise to the need for a corresponding strategy of corporate containment through mergers and acquisitions. Here, again, the rapid pace of technological change presented a *problem*, rather than an aid to the differential accumulation of dominant firms, and the solution was to subjugate technical change by buying up—rather than competing with—new firms.

The following analysis uses the growth of new firms as a quantitative proxy for these centrifugal forces, in order to show their two-fold impact. First, the rise in new firms through the mid-1990s and early 2000s undermined certainty about the relative power of Dominant Semiconductor Capital, quantitatively expressed as greater volatility in capitalization. Second, the new firm growth created a pool of new targets for acquisition, and an external breadth wave of mergers and acquisitions began around the same time. At first, the pace of new firms' emergence ran ahead of dominant firms' mergers and acquisitions activity. By the 2010s, the trend reversed, and the increasing pace and size of M&A activity significantly reduced the overall number of firms. The result was a massive increase in the differential earnings of Dominant Semiconductor Capital and a reduction in the volatility of Dominant Semiconductor Capital's capitalization.

[16] These forces also worked in parallel and combination. For instance, Apple contracted eSilicon, the first "fabless ASIC" firm, to supply the system chip for the original iPod (41).

Figure 5. Annual volume of acquisitions by monetary value, semiconductor manufacturing

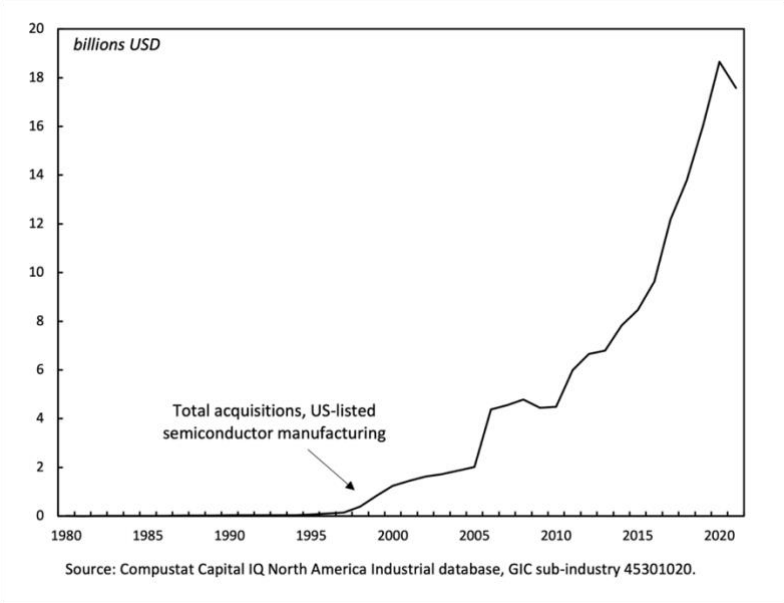


Figure 5 illustrates how dramatic this 'breadth' wave was. Acquisitions were negligible during the 1970s and 1980s but, starting in the mid-1990s, rose to several billion dollars a year. Though M&A activity slowed somewhat in the aftermath of the dot-com bust and again during the financial crisis of 2008-09, the sector has followed a path of more or less exponential growth in mergers and acquisitions activity.

'Industrial' growth and uncertainty

The rise of new firms is, in effect, green-field investment – by increasing the size of the overall earnings pie, it undermines the differential control over production held by Dominant Semiconductor Capital. This can be shown quantitatively in multiple ways.

Figure 6. Sector size and Dominant Semiconductor Capital's share of the revenue

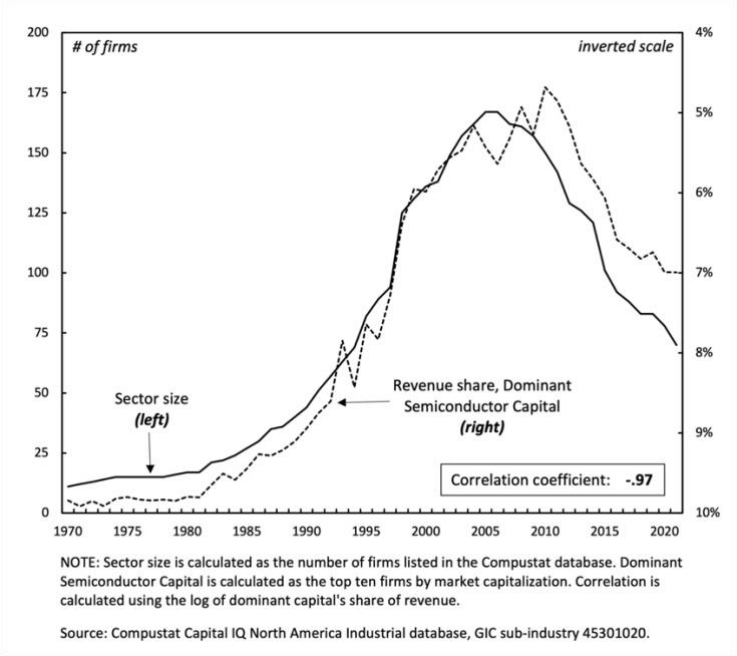
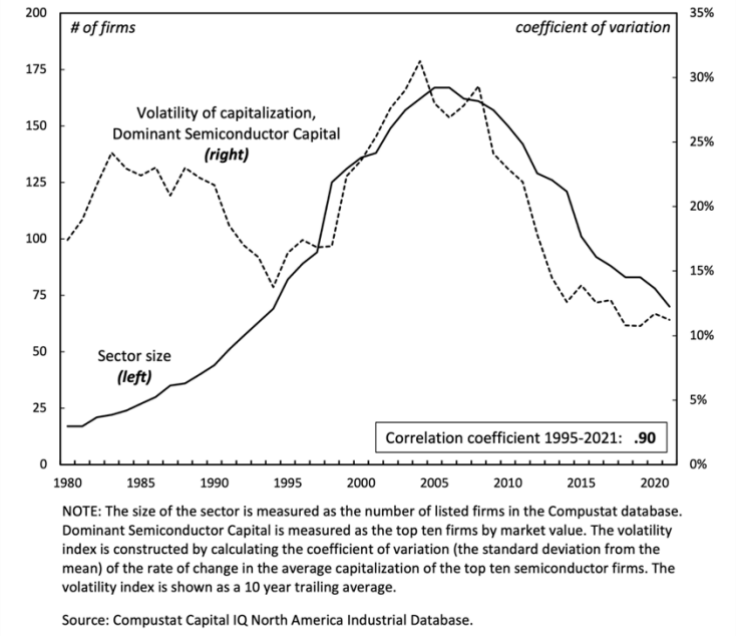


Figure 6 shows the sector size, measured as the number of semiconductor manufacturing firms listed in the Compustat database and the share of total revenue for the sector held by Dominant Semiconductor Capital. As the number of firms increases, the percentage of total sales captured by Dominant Semiconductor Capital shrinks. The tight negative correlation between the two series (-.97, note the inverted right scale) suggests that the sector's overall growth undermined the share of revenue of Dominant Semiconductor Capital, threatening their relative power. It also illustrates the danger for dominant capital of 'too much' green-field investment. As total production expands, Dominant Semiconductor Capital must grow *faster* than the average to maintain the same revenue share. However, if the ability to retain a differential profit *margin* is reliant on tightly restraining the expansion of production, any increase in production to keep up with the sector's growth will tend to work against this strategy.

Figure 7. Sector size and capitalization volatility, Dominant Semiconductor Capital



NOTE: The size of the sector is measured as the number of listed firms in the Compustat database. Dominant Semiconductor Capital is measured as the top ten firms by market value. The volatility index is constructed by calculating the coefficient of variation (the standard deviation from the mean) of the rate of change in the average capitalization of the top ten semiconductor firms. The volatility index is shown as a 10 year trailing average.

Source: Compustat Capital IQ North America Industrial Database.

Figure 7 confirms the destabilizing effect of new firm growth on the power of Dominant Semiconductor Capital using another way of measuring power: capitalization. According to capital as power theory, capitalization represents a symbolic estimation of a firm's ability to generate risk-adjusted earnings in the future. Capitalization is forward-looking and, in principle, a universal process. Nitzan and Bichler write: "if it generates earning expectations, it must have a price, and the algorithm that gives future earnings a price is capitalization" (Nitzan and Bichler 2009, 158). Because the future is inherently uncertain, and since capitalization is a future-oriented process, the volatility of a firm's capitalization can be understood as a quantitative proxy for measuring capitalists' level of certainty or uncertainty about their estimations of the future. In short, all else remaining the same, the less *certain* capitalists are of those future expectations, the more change or volatility one might see in a firm's capitalization.

Figure 7 shows the relationship between the size of the sector, measured as the total number of listed firms in the Compustat database, and the volatility of Dominant Semiconductor Capital's capitalization. Volatility here is calculated as the coefficient of variation of the rate of change in the average market value of Dominant Semiconductor Capital.[17] The correlation between new firm growth and the volatility

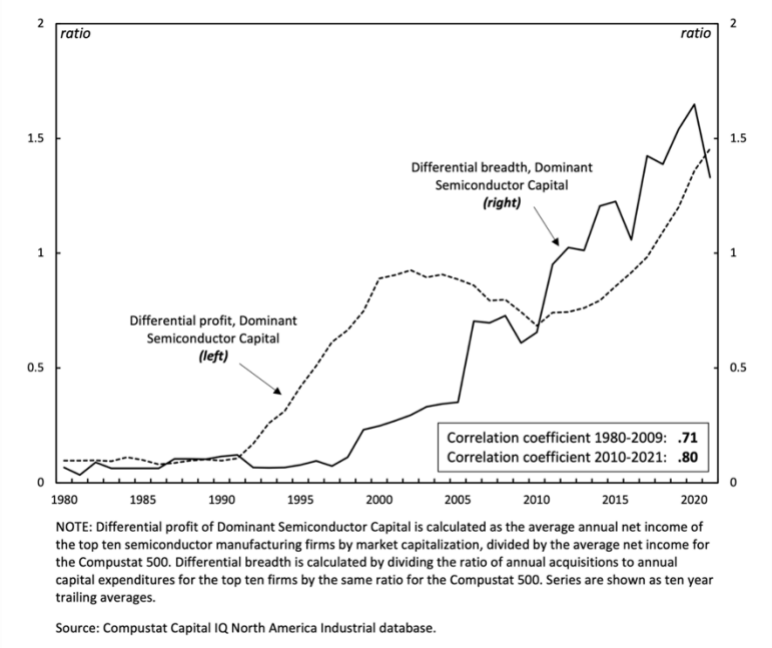
[17] The coefficient of variation is measured as the standard deviation from the sample's mean.

of Dominant Semiconductor Capital's capitalization between 1995-2021 (.90) suggests that as the sector expanded through the 1990s, the power of Dominant Semiconductor Capital became more *uncertain*. This pattern makes intuitive sense, as the growth of the number of firms is simply the growth in the number of potential challengers or competitors. With each new firm, the future power of current dominant firms is potentially diminished by an unknown degree. Therefore, a larger field of play is inherently more dynamic and riskier. The rise in volatility suggests that growth in the sector outpaced dominant firms' ability to buy up competitors, increasing uncertainty around Dominant Semiconductor Capital's differential power. By 2010, the trend had reversed: as the number of firms in the sector decreased and M&A activity skyrocketed, Dominant Semiconductor Capital's capitalization (read: estimated relative future earning capacity) became less volatile.

Differential buy-to-build

Figure 8 compares the differential buy-to-build ratio and differential profitability of Dominant Semiconductor Capital. Differential buy-to-build represents the emphasis of dominant semiconductor firms on investment in M&A over the expansion of productive capacity, relative to a benchmark.[18] Differential profitability is measured here as a ratio of the average net income of Dominant Semiconductor Capital to the average net income of a Compustat 500 firm.

Figure 8. Differential buy-to-build and differential profitability, Dominant Semiconductor Capital



Overall, there is a strong correlation between differential buy-to-build and differential profitability within Dominant Semiconductor Capital, particularly during 2010-2021 (.80). Up until the mid-1990s, both differential buy-to-build and differential profitability remained fairly level, possibly because of the lack of takeover targets in the semiconductor business. Starting from the end of the 1990s, differential buy-to-build rose steadily, while differential profitability initially rose, then moved sideways/fell through the 2000s. After 2010, both differential profitability and differential buy-to-build rose sharply, suggesting that

[18] For more on the 'buy-to-build' indicator, see the Appendix at the end of the paper.

the rapid gains of Dominant Semiconductor Capital during this period were caused at least in part by their differential pace of acquisitions.

If the creation of new ideas represents one pole of technological change within the semiconductor business, then the physical manufacture of the products of those ideas represents the other. On the one hand, the rapid pace of change in semiconductor development, combined with the periodic differential *de*cumulation of Dominant Semiconductor Capital, suggests a certain degree of failure to profitably limit the creation and diffusion of new ideas. On the other hand, the regular appearance of 'shortages' suggests that dominant semiconductor firms have primarily relied on the strategic limitation of the *physical manufacture* of chips for differential accumulation. The evidence in this section indicates that this strategy may have shifted as new design, IP, and other fabless firms emerged as targets for mergers and acquisitions. It also confirms the capital as power thesis that, under certain conditions, M&A activity is a reliable way to accumulate differentially.

A return to depth?

The current trade conflict with China in some ways parallels the conflict with Japan in the 1980s, not least because the semiconductor business appears to be implicated at the highest levels. While it is still a few 'production nodes' behind US and Taiwanese semiconductor firms, the Chinese ruling elite is pouring resources into achieving technological parity, if not supremacy, with the other technological superpowers (Capri 2020, 30). China is already the largest purchaser of semiconductors globally, and its technology sector is highly integrated with foreign capital (23). The combination of policies designed to encourage foreign investment and the desire to gain access to China's rapidly growing technology sectors have fueled a steady migration of semiconductor business into the country (Chu 2013, 188). A 2020 Hinrich Foundation report lists 22 semiconductor firms' joint ventures with China between 2014 and 2018, while "in China's computer sector, foreign-invested enterprises accounted for 59 percent of industry assets and 57 percent of industry profits in the manufacture of computers" in 2013 (Capri 2020, 34).

At the same time, the unleashing of industry under China's industrialization policy has also sparked fears of 'overproduction'. For instance, Capri estimates that "the possibility of an over-supply of NAND and DRAM chips would seem likely, at some point, which would drive down global market prices" (2020, 29). In other words, China appears to be pursuing a similar strategy as Japan did in the 1980s – a strategy of green-field investment that some worry will undermine the differential profitability of other firms. As Capri notes, "none of this bodes well for the world's existing players" (29).

It is possible that the trade war is making it increasingly difficult to continue the integration of Chinese firms into an increasingly global Dominant Semiconductor Capital. Starting in 2016, the Committee on Foreign Investment in the US (CFIUS) began closely scrutinizing M&A activity between the US and China and shut down several high-profile cases (46-48). Most notably, former President Donald Trump blocked a $117 billion acquisition of semiconductor manufacturing firm Qualcomm by (at the time) Singapore-based Broadcom – a deal which would have been one of the largest in history. Does this point to an end to the breadth wave of the 2010s? Since 2020, the world has been gripped by a global chip 'shortage'; in turn, chipmakers have booked record profits. Whatever the future holds for the US and China, Dominant Semiconductor Capital will likely continue to play a starring role.

7. Conclusion

Semiconductor technology and semiconductor firms do not exist in a vacuum – they are enmeshed in complex ways with social forces outside the control of any dominant group. The interests of the semiconductor business cannot wholly escape the influence of government and military agencies, computer and consumer electronics producers, or the tastes and fashions of computing culture at large. Neither are these institutions, individuals and dynamics easily separable from the sphere of semiconductor 'business' interests. From a capital as power perspective, the stock price of publicly traded semiconductor firms capitalizes the power of global intellectual property rights, Chinese government policy, innovations in the computing industry, worker migration and everything else that may bear on the future profitability of the company. Each of these is, in turn, shaped by the logic of differential accumulation.

That said, it is worth returning to Lewis Mumford's warning about equating all technological advancement with the development of human potential and well-being. His warning was this:

> While any new technical device may increase the range of human freedom, it does so only if the human beneficiaries are at liberty to accept it, modify it, or reject it: to use it where and when and how it suits their own purposes, in quantities that conform to those purposes. (Mumford 1971, 185).

It is not difficult to find examples of how computing technology has failed us in these respects. To name a few: the military use of semiconductors in ever more sophisticated and lethal weapons systems has undoubtedly made the world a more dangerous place; cheap computing power has led to the rise of socially and ecologically destructive practices like digital addiction and cryptocurrency mining; while the ownership of mobile phones—devices designed for the widespread and invasive surveillance of populations—is now a precondition for obtaining even the most basic social necessities of life. In this context, in the words of Mumford, the pressure to "forego all modes of activity except those that call for the unremitting use of the 'machine' or its products'" is perhaps itself a form of social sabotage (329). Thus, the conclusion of this paper is *not* that technological progress, as it stands today, might simply be accelerated by removing the profit motive from industrial organization. The solution to restraining business sabotage is not to allow the full expansionary dynamics of technological change to run independently of human and ecological needs, but to assert greater democratic control over the direction of technical change, including restraining technological advancement where it threatens our collective wellbeing.

Appendix: data and methods

In examining quantitative evidence regarding the semiconductor business, I draw on several concepts and procedures developed in the capital as power literature and add a couple of my own. My main source of empirical evidence is financial data relating to profit and investment by semiconductor firms. My sources are the Compustat Capital IQ financial database, accessed through Wharton Research Data Services (WRDS), and the financial database *Mergent Online*. Data are processed using the statistical tools of Microsoft Excel.

Measuring 'Dominant Semiconductor Capital'

The dominant capital concept is central to studying capital as power. While all capitalists are compelled to differentially accumulate to survive, our inquiry concerns those that succeed and play a dominant role in shaping society. While dominant capital can be understood as an objective social category, measuring the bounds of dominant capital is always partly a matter of interpretation and conjecture. On the one hand, one only has to think of blue-chip indexes like the Nasdaq or the S&P 500 to evoke the idea of dominant capital and the many ways one might define it empirically. Treated as a quantitative abstraction, it appears as an index of the most powerful groups in society, comprising a more or less integrated coalition of increasingly global firms, owners, governments, and other large organizations. On the other hand, any quantitative representation will always inevitably fall short of describing the complexity and dynamism of the underlying reality.

Within the capital as power literature, dominant capital is usually measured as the numerical or statistical grouping of the top firms by revenue or market value (capitalization). It can be defined as either a fixed numerical group (for example, the top ten, fifteen, or one hundred) or a percentage (the top one, five, or ten percent, and so on.) of firms within a given social space. I follow this practice, noting again that such a measure is only approximate and does not necessarily represent a coherent social unit by other criteria. Moreover, because the social space is dynamic and overdetermined by many social relations, the quantitative analysis is accompanied by an attempt to qualitatively characterize the same power relations in their historical and processual specificity. Firms comprising dominant capital may struggle among themselves as much as they protect specific collective interests; the capitalization of a given firm may be underpinned by racial violence, national rivalries, and gender inequality; and resistance to capitalist power—often subterranean and unpredictable—is always present, though difficult to quantify. Nevertheless, if capital, measured differentially, represents capitalists' quantified estimates of those forces, and to a large extent, these estimations guide capitalist behaviour, then there are likely to emerge quantitative patterns that align with the qualitative reality.

In this paper, the concept 'Dominant Semiconductor Capital' is used as an approximation for the most powerful US-listed semiconductor manufacturing firms. To construct this empirical measure, I started with the entire North America Compustat Industrial database from 1970-2020. This database comprises the largest public firms listed on North American stock exchanges. I then isolated all firms with the GIC industry code 45301020, which includes only semiconductor manufacturing firms. Firms that produce semiconductors, but either do so solely for in-house use (for instance, IBM) or as unlisted subsidiaries of a larger corporation (for instance, Samsung) are excluded from this list. In the first case, these producers to not share the same business interests as merchant producers. They do not profit from the production of semiconductors directly, but only from the sale of goods containing semiconductors. As such, I assume that they do not necessarily share an interest in raising the price of semiconductors through strategic sabotage. In the second case, the issue is more empirical than theoretical. It is often difficult, if not impossible, to determine from the available financial data what proportion of a firm's total

revenue or income is derived from semiconductor manufacture. In the end, I decided that the risk of contaminating the analysis with unrelated financial data was greater than the benefit of trying to include these firms.

In addition, some of the firms listed in the Compustat database are not US-based companies. However, my goal is not to construct a US-only database but rather an approximate index of global semiconductor firms that can be easily measured. From the 1980s onwards, Dominant Semiconductor Capital increasingly organized itself globally, rather than nationally, even if the power of national governments remains a necessary part of the differential power of those firms, and most of the leading members of this groups are listed in the US. Significant omissions in this context, due to the scope of the project and the availability of data from these regions, are several Japanese and Chinese firms which would make the dataset more representative of the global state of the business.[19] While this is an unfortunate limitation, such additions must wait for a later project.

After constructing the 45301020 (semiconductor firm) database, I created two Dominant Semiconductor Capital datasets: one comprising the top ten firms by market capitalization and the other the top 10% of firms by market capitalization (reselected annually).[20] The results between the two do not differ greatly, and so in the interest of readability, I presented my analysis using only the first method.

Constructing the benchmark

To create a benchmark against which to measure the differential accumulation of Dominant Semiconductor Capital, I constructed a set of the largest 500 firms by capitalization within the entire Compustat North America Industrial database (reselected annually). I then calculated the weighted average of each relevant data point – revenue, profit, capital expenditures, acquisitions, etc. This benchmark dataset is meant to represent the average large firm, against which dominant semiconductor firms might theoretically judge their own accumulation, similar to a measure like the S&P 500. To calculate the weighted averages, I divided the total revenue (or profit, capital expenditures, etc.) by the number of firms, and used these averages to calculate derived measures, like average return on equity. The benefit of this approach (as opposed to using unweighted averages) is that in a weighted average, the size of individual firms determines the size of their effect on the calculation of the average. Because I am focusing on the behavior of large firms under the assumption that larger firms are large *because* of their distinguishing behavior, giving equal weight to smaller and larger firms could potentially be misleading as to the content and effects of that behavior.

Measuring 'shortages'

A central theoretical argument of capital as power research is that price changes are manifestations of social power and the ability to use it strategically. In analyzing the role of 'shortages' in the semiconductor business, the empirical object is two-fold. First, we must determine if there is a relation between 'shortages', price changes and profits; Second, we must evaluate what qualitative power relations underpin these quantitative relationships.

[19] In the case of Japan, and for the greater part of the period analyzed, most semiconductor manufacturing firms were also subsidiaries of larger consumer electronics conglomerates.

[20] The list of firms was recalculated for each year based on that year's leading firms.

How do we determine what constitutes a shortage? Unfortunately, this question may, in part, have to remain unanswered. Principally, the term 'shortage' is problematic because it is conventionally defined in terms of neoclassical economic theory. According to neoclassical theory, a shortage is a situation in which, at a given price, the quantity demanded exceeds the quantity supplied. The problem with this formulation is that neoclassical demand and supply denote the desires of buyers and sellers, and these desires cannot be empirically observed. Since no one can tell the difference between these unobservable magnitudes, there is no way to tell the extent of the resulting 'shortage'. Even if a shortage is described in 'practical' terms —for instance, when a greater number of orders is received than can be physically manufactured within a given time frame in a given factory—problems of measurement and definition still arise. For instance, how does one know whether chip users intend to buy the chips they order? During a 'shortage', chip users can place multiple orders at different firms and then later cancel some of them, or buy more chips than necessary, and then sell the surplus (Flamm 1996, 233). How does one determine what is real and what is fake 'demand' from these possibilities?

Similarly, how does one determine a chip factory's 'true' technical capacity? Factories rarely work at 'full' capacity and available capacity utilization measurements do not reflect *technical* capacity utilization but *profitable* capacity utilization (Nitzan and Bichler 2009, 234). A firm's need to achieve a specific rate of return shapes its perceptions about the correct production level, apart from considerations of the needs of chip users (which, of course, may be just as difficult to measure). Neoclassical economics assumes that full technical utilization and profitable full utilization are equivalent. However, even concrete attempts to define 'shortage' still face theoretical and methodological difficulties if this assumption is not granted. Thus, another methodology is needed to examine the role of 'shortages' in the semiconductor business.

One empirically observable phenomenon is the *perception* that there is a shortage. The perception of a shortage is relevant because if the actual existence of a shortage is difficult, if not impossible, to verify, then it is *only* the perception that in the end, *justifies* the raising of prices. Justification is necessary because price increases tend to redistribute income from buyers to sellers (provided the level of quantities exchanged does not drop so low that the redistribution is reversed), and thus rising prices tend to meet resistance. Resistance can be countered by having a good *reason* for raising prices, preferably one that places the blame for higher prices outside the seller's control. The presumed existence of a shortage fulfills this social requirement, whether or not there is a 'real' shortage. It is not necessarily a question of the seller 'lying' about a shortage – it can simply be a matter of reducing production or restraining investment in new production capacity until a shortage becomes 'inevitable'. By measuring the perception of shortages, we can examine its correlation with prices while avoiding the theoretical and empirical problems of measuring the shortage itself, and without losing sight of the relationship between shortages and prices as one of justification in the face of resistance.

In short, the methodological problem of measuring 'shortages' is provisionally solved by looking only at the perception of shortages. The paper does this by using public mentions of semiconductor shortages through news publications as a proxy for the existence of a general perception. Using the *Nexis Uni* online database, I made a targeted search to identify all newspaper articles between 1980 and 2020 reporting on a semiconductor shortage in either the headline or the body of the text.[21] A list of the headlines and links to the articles were downloaded into Microsoft Excel and filtered to find all *headlines* mentioning chip shortages, for a total of 83 articles spanning April 1984 to December 2020. For expediency, results were not counted for 2021, because the first few pages of results returned over one

[21] The Boolean expression used in the search was < ("semiconductor" OR "chip") AND "shortage">, and the search was limited to English language publications.

hundred headlines mentioning chip shortages. The annual number of mentions is not relevant to the study – one or more mentions in a year is counted as the existence of the perception of a chip shortage.

It is plausible that greater mentions during a given year indicate either a more widespread perception of shortage, a perception of greater seriousness of shortage, or both. However, this study does not attempt to quantitatively define either widespread-ness or seriousness, labelling years only as having a perception of a shortage or not. Finally, the focus on the perception of shortages over 'real' shortages does not exclude the possibility of a technical shortage. If it is the perception that a shortage is occurring that ultimately justifies the price increase, whether or not a shortage is actually occurring is irrelevant to our inquiry.

Measuring 'breadth'

To analyze the breadth strategies of Dominant Semiconductor Capital, I constructed 'buy-to-build' indicators for Dominant Semiconductor Capital and the Compustat 500. This indicator is intended to approximate the relative emphasis on external (M&A) over internal (green-field) breadth. It is calculated as the ratio of acquisition expenditures to new capital expenditures (creation of new plant, property, and equipment). This indicator, first conceived by Nitzan and Bichler, is a novel measurement related directly to capital as power research.[22] Whereas green-field investment tends to become unruly and undermine differential profit, mergers and acquisitions increase firms' relative size without increasing the total run of production (Nitzan and Bichler 2009, 335). Thus, the logic behind the measurement is that, over time, firms pursuing a breadth strategy will devote an increasing share of investment to acquisitions, and dominant firms, *in particular,* will tend to exhibit a higher ratio of acquisitions to green-field investment than the average firm over time. In the above section "Centrifugal forces" I further develop this concept by measuring the differential buy-to-build ratio of Dominant Semiconductor Capital compared to the Compustat 500 average.

[22] For more on the buy-to-build indicator, see Nitzan and Bichler 2009, Ch. 15; and Francis 2013.

References

Bichler, Shimshon and Jonathan Nitzan. 2013. "Francis' Buy-to-Build Estimates for Britain and the United States: A Comment." *Review of Capital as Power* 1, no. 1: 73-78.

Bichler, Shimshon and Jonathan Nitzan. 2021. "The 1-2-3 Toolbox of Mainstream Economics: Promising Everything, Delivering Nothing." *Real-World Economics Review* no. 98 (December): 23-48.

Brown, Clair and Greg Linden. *Chips and Change: How Crisis Reshapes the Semiconductor Industry*. Cambridge: MIT Press.

Capri, Alex. 2020. "Semiconductors at the Heart of the US-China Tech War: How a New Era of Techno-Nationalism is Shaking up Semiconductor Value Chains." The Hinrich Foundation.

Chu, Ming-chin Monique. 2013. *The East Asian Computer Chip War*. New York: Routledge.

Flamm, Kenneth. 1996. *Mismanaged Trade? Strategic Policy in the Semiconductor Industry*. Washington, DC: Brookings Institution.

Flannery, Russell. 2022. "Global Chip Supply-Demand Imbalance Is Spurring Industry Change: NXP CEO Kurt Sievers." Forbes.com, May 25, 2022. https://www.forbes.com/sites/russellflannery/2022/05/25/global-chip-supply-demand-imbalance-is-spurring-industry-change-nxp-ceo-kurt-sievers/?sh=24421ac222f5.

Francis, Joseph A. 2013. "The Buy-to-Build Indicator: New Estimates for Britain and the United States." *Review of Capital as Power*, Vol. 1, No. 1, pp. 63-72.

Grindley, Peter, and David Teece. 1997. "Managing Intellectual Capital: Licensing and Cross-Licensing in Semiconductors and Electronics." *California Management Review* 39, no. 2: 8–41.

Hart, David M. 2001. "Antitrust and Technological Innovation in the US: Ideas, Institutions, Decisions, and Impacts, 1890–2000." *Research Policy* 30, no. 6: 923–936.

Irwin, Douglas. 1996a. "The U.S.-Japan Semiconductor Trade Conflict." In *The Political Economy of Trade Protection*, edited by Anne Krueger, 5-14. Chicago: University of Chicago Press.

Irwin, Douglas. 1996b. "Trade Policies and the Semiconductor Industry." In *The Political Economy of American Trade Policy*, edited by Anne Krueger, 11-72. Chicago: University of Chicago Press.

Kelly, Leanna. 2022. "Global Semiconductor Chip Shortage Update." Investment U, March 14, 2022. https://investmentu.com/chip-shortage-update/.

Malerba, Franco. 1985. *The Semiconductor Business*. London: Franco Pinter.

Malone, Michael. 2014. *The Intel Trinity: How Robert Noyce, Gordon Moore, and Andy Grove Built the World's Most Important Company.* New York: Harper Business.

Martin, Ulf. 2019. "The Autocatalytic Sprawl of Pseudorational Mastery." *Review of Capital as Power* 1, no.4: 1-30.

McClean, Bill. 2011. "Research Bulletin: Tracking the Top 10 Semiconductor Sales Leaders Over 26 Years." *IC Insights*.

Mumford, Lewis. 1971. *The Pentagon of Power*. London: Secker and Warburg.

Nenni, Daniel, and Paul McLellan. 2013. *Fabless: The Transformation of the Semiconductor Industry*. CreateSpace Independent Publishing Platform.

Nitzan, Jonathan. 1992. "Inflation as Restructuring: A Theoretical and Empirical Account of the U.S. Experience." Unpublished PhD dissertation.

Nitzan, Jonathan, and Shimshon Bichler. 2002. *The Global Political Economy of Israel*. London: Pluto.

Nitzan, Jonathan, and Shimshon Bichler. 2009. *Capital as Power: A Study of Order and Creorder. RIPE Series in Global Political Economy*. New York and London: Routledge.

Noble, David. 1979. *America by Design: Science, Technology, and the Rise of Corporate Capitalism*. Oxford University Press.

Tilton, John. 1971. *The International Diffusion of Technology: The Case of Semiconductors*. Washington D.C.: The Brookings Institute.

Veblen, Thorstein. 1908. "On the Nature of Capital: Investment, Intangible Assets, and the Pecuniary Magnate." *Quarterly Journal of Economics* XXII (1, November): 104-136.

White, Edward, and Kathrin Hille. 2021. "Samsung warns of deepening semiconductor shortage." *Financial Times*, March 17, 2021. https://www.ft.com/content/7e5e2d0f-3aa4-412b-898c-15acda4adb2e.

Author contact: cjmoure@yahoo.com

SUGGESTED CITATION:
Christopher Mouré, "Technological change and strategic sabotage: A capital as power analysis of the US semiconductor business", *real-world economics review*, issue no. 103, 31 March 2023, pp. 26-55, http://www.paecon.net/PAEReview/issue103/Moure

Do copyrights and paywalls on academic journals violate the US Constitution?[1]

Spencer Graves [Founder, EffectiveDefense.org, Kansas City, Missouri USA]

Abstract
This article argues that paywalls and copyrights on academic journals are obstacles to the progress of science and the useful arts, therby violating the copyright clause of the Constitution of the United States. They also detract from rather than promote the general welfare, violating the purpose of the US Constitution.

1. Introduction

What percent of authors of articles in refereed academic journals are motivated to write for the royalties they expect to receive from publication?

More specifically, do *any* refereed academic journals offer substantive payment for articles published? In rare cases, the author of this article has seen prizes for, e.g., best paper of the year. However, to the extent that this author's experiences are typical of academic publishing, no reasonable human would submit an article to a refereed academic journal expecting substantive income derived from copyright royalties. Instead, researchers write to be read and cited, to contribute to culture and the shared body of knowledge of humanity, and to build their reputations. Many also write in part because the criteria by which they are hired and promoted are based partly on publication records.

In contrast, for many refereed academic journals, authors are required to assign the copyright to the journal as "a work made for hire", even though it was *not* "made for hire." Prior to the Internet, that was justified to cover the costs of printing and distribution. Those days are past.

[1] A preliminary version of this is available on Wikiversity, "Do copyrights and paywalls on academic journals violate the US Constitution?" (https://en.wikiversity.org/wiki/Do_copyrights_and_paywalls_on_academic_journals_violate_the_US_Constitution%3F), accessed 11 November 2022.

2. Copyright Clause and Preamble of the US Constitution

The Copyright Clause of the US Constitution says, "The Congress shall have Power ... To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries".[2] Moreover, the Preamble to the US Constitution says its purpose includes promoting "the general welfare".
People who submit articles to refereed academic journals do so to be read and to build their reputations. To the extent that this is accurate:

- Paywalls on academic journals are obstacles to "the progress of science and the useful arts". They are obstacles to promoting the general welfare, in apparent violation of the US Constitution.

3. Why is this part of US and international law?

Major media have a conflict of interest in honestly reporting on anything relating to copyright law. McChesney (2004, p. 233) wrote that the US "Congress initially called for copyright terms of fourteen years. From the beginning, prominent publishers such as Noah Webster worked to expand the length and terms of copyright. ... But the public benefitted tremendously from having limited copyright ... . The balance between copyright and public domain began to change dramatically in the twentieth century with the rise of the corporate media system", with changes in copyright law underreported and approved by legislators, many and probably most of whom received "campaign contributions"[3] from such media. Most of those legislators also doubtlessly knew that it would likely be political suicide to oppose the mainstream media.[4] The Electronic Frontier Foundation recently asked supporters to "Stop copyright creep", noting that "an unconstitutional proposal and [bill] that no one in the public had ever seen before" made it into the 2020 year-end must-pass omnibus bill.[5]

One important US Supreme Court decision in this regard is *Eldred v. Ashcroft*. That lawsuit challenged the Sonny Bono Copyright Term Extension Act of 1998, which added 20 years to the terms of copyrights, including existing copyrights. That act has also been called the "Mickey Mouse protection act", because it extended from 1998 to 2018 the date at which the first commercially successful Mickey Mouse cartoon, released in 1928, entered the public domain. It is difficult to understand how extending the term of existing copyrights *retroactively* could "promote the progress of science and the useful arts". However, the Supreme Court insisted that Congress had the authority to make that judgment. Lawrence Lessig, the lead attorney for the plaintiffs, said he lost that case, because he focused too much on law and too little on economics. Lessig wrote, "Of all the creative work produced by humans anywhere, a tiny fraction has continuing commercial value. ... But even for that tiny fraction, the actual time during which the creative work has a commercial life is extremely short. Most books go out of print within one

[2] US Constitution, Article I, Section 8, Clause 8. See also the discussion in the Wikipedia article on "Copyright Clause".

[3] Political campaign contributions are not "bribes" under US law unless there is an explicit quid pro quo. However, their corrupting influence on US law has been extensively documented, though not widely reported, perhaps because the major media are a major recipient of the largess. See, e.g., Teachout (2014), Lessig (2011), Johnston (2007). Stern (1992) documents multiple cases where major campaign contributors got over $1,000 return for each $1 invested in political campaigns.

[4] e.g., McChesney (2004). See also Rolnik et al. (2019).

[5] Electronic Frontier Foundation (2022).

year. The same is true of music and film."[6] Might it have been easier for Lessig to have won that case if relevant empirical economics research had been more substantive and more available? For example, what is the implied discount rate for investments in copyrights, and how might that be impacted by different copyright terms?[7]

Similar issues are raised by the current Publishers' lawsuit against the digital lending program of the Internet Archive, where digitized copies of books in that online library can be checked out for an hour at a time from anywhere in the world with an Internet connection.[8]

In *Free Culture*, Lessig argued that a major obstacle to the evolution of culture is the provisions of copyright law requiring payment of royalties for "derivative works". A natural experiment in this regard compares the "Manga" comics industry in Japan with the rest of the world. Lessig wrote that in Japan, "Some 40 percent of publications are comics". He claims that a similar comics industry is not feasible in the US and many other countries, because most of those comics might be considered "derivative works" of previous works, requiring prohibitive copyright permissions prior to publication. Even the first successful Mickey Mouse cartoon might not have appeared under current copyright law, because it might have been challenged as a "derivative work" of an earlier Buster Keaton film. The written law in Japan is similar to the US in this regard, but the Japanese largely ignore the provisions for "derivative works".[9]

Returning to the primary question of this article, are there *other* justifications for paywalls and copyrights held by academic journals?

In particular, are copyrights and paywalls on academic journals *accidents*? Are they due to failures of the US Congress to fully consider the ramifications of what they've enacted? Or are they monuments to political corruption?

[6] Lessig (2004b). This book documents how excessive copyright terms are obstacles to documenting and understanding of the evolution of culture. Many old films are physically decaying beyond use in private archives generally unaccessible, even to scholars. This practice protects copyright revenues from the very few films that still have value, far beyond the planning horizon of any rational investor.

[7] "Pricing risky corporate bonds" seems like a similar issue, discussed by Baaquie and Karim (2022). Studies of the political economy of copyright law could benefit from similarly considering intellectual property law more generally, especially patent law. For a brief overview of the impact of patent law on public health, especially infectious diseases, see Graves and Samuelson (2022).

[8] See the section on "Publishers' lawsuit" and references cited therein in the Wikipedia article on Internet Archive, accessed 12 November 2022.

[9] Lessig (2004a, p. 24). See also the Wikipedia article on "*Free Culture* (book)".

4. What to do?

To the extent that copyrights and paywalls on academic journals are obstacles to "the progress of science and the useful arts", there are things that individual researchers, academic administrators, and the public can do to help overcome these obstacles:

- **Researchers** can submit their work only to open-access journals and refuse to submit their work to journals that will put their work behind a paywall. (No one who wants to be cited wants their work behind a paywall if there is a reasonable alternative, because the paywall would likely reduce their audience.)

- **Administrators** managing research that produce articles for academic publications can insist that their researchers submit their work only to open-access journals. (Anyone wanting to build the reputation of their research wants their publications to be read. Paywalls and copyrights are obstacles to that.)

- **Citizens** should demand that their elected officials enact two reforms affecting copyrights:

    1. All government funded research should be freely available, not behind a paywall, and should either be in the public domain or with a license no more restrictive than the Creative Commons Attribution-ShareAlike (CC BY-SA) 4.0 International license.[10]

    2. Copyright law should be changed to *forbid* restrictive copyrights on "works for hire" when they are *not* actually written with a plausible expectation of receiving substantive income derived from copyright royalties. This would leave in place current practices for publications other than academic journals.

References

Baaquie, Belal Ehsan; and Karim, Muhammad Mahmudul (28 September 2022) "Pricing risky corporate bonds: An empirical study". *Journal of Futures Markets*.[11]

Electronic Frontier Foundation (1 November 2022) "Tell Congress to Keep Copyright Out of Must-Pass Legislation" (https://supporters.eff.org/civicrm/mailing/view?reset=1&id=11668), accessed 15 November 2022.

Graves, Spencer, and Samuelson, Douglas A. (2022) "Externalities, public goods, and infectious diseases", *Real-World Economics Review*, **99**: 25-56.

Johnston, David Cay (2007) *Free Lunch: How the Wealthiest Americans Enrich Themselves at Government Expense and Stick You With The Bill*, Penguin.

Lessig, Lawrence (2004a) *Free Culture*, Penguin.

Lessig, Lawrence (2004b) "How I Lost the Big One". *Legal Affairs*.

Lessig, Lawrence (2011) *Republic, Lost*, Hachette.

[10] Available from Creative Commons website (https://creativecommons.org/licenses/by-sa/4.0/).

[11] Full disclosure: The author has read only the freely available title and abstract of this article. The full article is behind a paywall. Present purposes seem to justify this citation without the expense in money and time needed to review the entire article.

McChesney, Robert W. (2004) *The Problem of the Media: U.S. Communication Politics in the 21st Century*, Monthly Review Press.

Rolnik, Guy; Cagé, Julia; Gans, Joshua; Goodman, Ellen P.; Knight, Brian G.; Prat, Andrea; and Schiffrin, Anya (1 July 2019) *Protecting Journalism in the Age of Digital Platforms* (PDF), Booth School of Business.

Stern, Philip M. (1992) *Still the best congress money can buy*, Regnery Gateway.

Teachout, Zephyr (2014) *Corruption in America: From Benjamin Franklin's snuffbox to Citizens United*, Harvard U. Pr.

Author contact: spencer.graves@effectivedefense.org

SUGGESTED CITATION:
Spencer Graves, "Do copyrights and paywalls on academic journals violate the US Constitution?", *real-world economics review*, issue no. 103, 31 March 2023, pp. 56-60,
http://www.paecon.net/PAEReview/issue103/Graves

You may post and read comments on this paper at http://rwer.wordpress.com/comments-on-rwer-issue-no-103/

real-world economics review, issue no. 103
subscribe for free

Mainstream economics – the poverty of fictional storytelling

Lars P. Syll [Malmö University, Sweden]

Introduction

Mainstream economists usually content themselves with stating something like "economics is what economists do" and "economists are those who deal with economics." No deeper philosophizing is considered necessary.

However, this is an unsustainable and not particularly enlightening attitude. Scientific, philosophical, and methodological analyses of economic science are both important and necessary. Methodological knowledge serves as a 'road map'. When we discover that we are not on the road, we must occasionally look at the map or model for scientific progress that we all carry with us, consciously or unconsciously. Economic theories *always* build -- consciously or unconsciously -- on methodological positions. The question is therefore not whether economists should deal with methodology or not, but rather *how* it should best be done. Methodological analysis is both desirable and inevitable in economic science. Not least, it can have a critical function by making economists aware that the fundamental flaws of economic theory may be due to the fact that the concepts, theories, and models they use are incompatible with the very object of investigation. The tools borrowed mainly from physics and mathematics were constructed to solve completely different tasks and problems and risk contributing to a non-correspondence between the structure of economic science and the structure of economic reality. This, in turn, can lead economists to dubious simplifications and generalizations. Predictions and explanations based on such unauthorized 'bridges' between theory and reality have limited value and reliability.

The starting point for my criticism of mainstream economics is critical realism, a scientific and methodological direction primarily developed by the philosopher of science Roy Bhaskar. In *A Realist Theory of Science*, Bhaskar (1997) laid the foundation for a scientific theoretical stance that primarily means that the basic ontological question -- how reality must be constituted for us to be able to know something about it -- is prioritized over the epistemological question of how knowledge is possible. Against *naive realism*, critical realism asserts that we must focus our attention primarily on the underlying causal mechanisms of reality rather than just studying empirically observable events.

Economic methodologists try to develop and apply criteria for evaluating different theories (coherence, correspondence, clarity, simplicity, etc.). They often differ in their views on what constitutes reliable sources of knowledge (observation, experiment, theoretical reflection), what theories should look like (descriptive, analytical), how they should be tested (econometrics, case studies, randomized controlled studies, etc.), and whether they should be mathematically formulated or not. But the most fundamental question economic methodologists try to answer is what the purpose of economic analysis is. Although opinions are many, it can be said that there are two different fundamental views. One is that the goal of

economic analysis is to better predict what happens in the economy and to develop better instruments for steering the economy in a certain direction. The other -- to which a critical realist methodology must be counted -- asserts that economic analysis must first and foremost aim to *understand* and *explain* how the economy works.

The inadequacy of instrumentalism

Realism as a methodological approach has been remarkably neglected or even unnoticed by mainstream economists. To the extent that it has been addressed, it is largely in connection with the debate over Milton Friedman's (Friedman, 1953) defense of the thesis that the realism of economic theory's assumptions is irrelevant. Against this background, it is perhaps not surprising that economists have recoiled from arguments presented from a realist perspective.

Friedman's instrumentalism is a particular version of empiricism according to which the goal of science is to formulate theories that can predict empirical events. Theories should therefore not be judged based on their explanatory power, but on how good instruments they are for delivering correct predictions. Theories and theoretical concepts should be understood as useful fictions that facilitate our predictions. If there is a 'real' world behind the theories we use, the instrumentalist is not concerned. Science cannot function as a bridge between the observable and the non-observable.

This 'fictionalist' perspective leads to the perception of the assumptions of economic theory as 'as if' assumptions, whose value lies in the fact that we can construct models that simplify complex contexts and make them analyzable and understandable.

For a scientific realist, Friedman's approach is unsatisfactory. If a theory proves to be useful, this is not a coincidence but depends on the nature of the object of knowledge. For a realist, it is necessary to try to find assumptions in theories that are practically relevant and that work and are consistent even in other contexts and with other known knowledge and practices. Theories should be *robust* and able to explain studied events by indicating what caused them, not just logically deducible.

Explanation and prediction of economic phenomena require theory construction, and seeking event correlations is therefore not enough. One must simply 'get beneath the surface' and see the underlying structures and generative mechanisms that explain the economic system.

Instead of explanations and predictions based on events and experiences, the primary object of investigation for critical realism is the *generative mechanisms* in the 'domain of reality' that cause events. The fact that economics has had such difficulty predicting events is not primarily because the theory would be incorrect, but because the underlying methodology assumes that the world consists of event regularities. Such regularities exist only if we assume that the world is *closed* -- which is a deeply unrealistic assumption.

Mainstream theory has its own solution to the problem -- it introduces a *ceteris paribus* clause. However, this is done at the expense of the fact that economics studies a theoretical and closed construction of the economy rather than the real economy, which is open. A paradoxical fact associated with mainstream theory is that it is said to be a theory of free individual choices in the market. But to predict events in such a model, choices must be 'closed' and preferences must be given and stable -- in short, the individual must be described as lacking real choices. They try to explain an event in an open social world (Y) from assumptions about what happens in a model that only applies in a closed world (X). But the explanation cannot be transferred from X to Y.

The atomistic fallacy

Methodological individualism -- explaining social institutions and change as a result of individual actors' actions and interactions -- has always had a strong position in mainstream economics.

But social and economic phenomena cannot be adequately explained by breaking them down into their smallest components or underlying strata without losing important emergent properties. Against the mainstream theory's reductionist view of 'macro explanations' as redundant once the micro level is mapped, critical realists object that social and economic contexts cannot be reduced to the individual level.

In economic theory, attempts are often made to build theory 'from the bottom up,' primarily expressed in the project of creating 'microfoundations' for macro theory (cf. Syll 2016). Sometimes this is done through careless aggregation, which is guilty of the so-called *atomistic fallacy*. The problem is that all the effects of microeconomic phenomena cannot be captured by simple aggregation. Although mainstream economic theory considers the mutual interdependencies of the parts, it often misses the more important emergent fact that the parts themselves can change and are partly constituted by the whole. The whole is not just a sum of parts but also something that affects and changes the nature and functions of the parts. From a critical realist view of social structures as real and independent entities, the entire micro-based project does not seem particularly fruitful.

When the model becomes the message

Economics is a science that relies heavily on the use of models. In economics textbooks, the use of models is generally presented as simplified descriptions of reality with which one constructs a kind of thought experiment and tests various hypotheses. The simple supply and demand model is a typical example. Mainstream economists admit -- usually -- that reality does not look like economic models, but that models have a justification as a kind of simplified *benchmark* with which the economist can describe and understand how economic mechanisms work.

But if the model does not reflect reality, *how* can we benefit from it? Is 'simplicity' and the ability to lead to 'clear conclusions' the model's most relevant aspect? How can we be sure that the model abstracts the 'essential' features? There is something deceptive about the evasive justification. It gives the *appearance* of qualifying the use of models in economics, but at the same time, it leaves open the question of how and to what extent models can contribute to our understanding of economic reality.

Sometimes it is argued that the use of formal models is good because they can guarantee logical consistency, force explicit assumptions, enable a concentrated representation, and facilitate communication among researchers. This may be true, but one should not forget that these advantages often lead to simplicity and mathematical elegance replacing explanatory power, that the models give rise to counterintuitive and paradoxical results, and that the desire for formalization can become an end in itself.

Model simplifications can be *harmless* or *restrictive* (depending on whether they affect deductions made within the theory in a decisive way or not). If the simplifications are of the restrictive type, the theory/model cannot be adequately used to explain or predict real events. If the restrictive assumption is removed, we cannot be sure that the relationships established in the theory/model remain.

The Iron Cage of Mathematics

Since the mid-20th century, mainstream economics has increasingly meant examining the world using the tools of mathematics. However, instead of uncritically adopting a mathematical representation form, economists should ask themselves what conditions the real processes and objects must meet for mathematical representations of them to be relevant. When the economic relationships or objects we are measuring or representing undergo qualitative changes, we cannot, just because a mathematical representation for 'tractability reasons' often requires it, model them as if they were given constants.

The precision and 'clarity' that mathematical language brings are no guarantee that the models are 'true.' Like any other models, they must be confronted with empirical observations and theory to determine whether they are adequate representations of reality.

Mathematics *can* be an excellent tool for model construction. But it must not become an end in itself. The fact that economic science is more quantitative than other social sciences is partly due to the fact that its objects of study are largely naturally quantitative (money, accounts, salaries, profits, etc.). However, this cannot be a defense for driving the mathematization trend *in absurdum* or for failing to ask what the mathematical models and quantitative measures are models of and measures of.

An exaggerated emphasis on the strength of mathematics and quantitative methods is typical of mainstream economists' deductivist approach. The dominance of the deductivist method in economic science is largely due to its suitability for mathematical modeling and its dominance in natural science (cf. Syll 2023). Since the end of the 19th century, neoclassical economics has seen natural science as a model for developing a supposedly more 'scientific' economic discipline.

The price paid for the mathematization of economics is that the subject has had to abandon the real world in order to focus scientific aspirations on proving things about imaginary worlds. Instead of accepting that a lower degree of certainty is inevitable, one engages in axiomatic and rationalistic model constructions that enable secure knowledge. If the goal is knowledge of the real world, the value of these is at best unclear.

No, there is nothing wrong with mathematics per se. No, there is nothing wrong with applying mathematics to economics. Mathematics is one valuable tool among other valuable tools for understanding and explaining things in economics. What is, however, totally wrong, are the utterly simplistic beliefs that

- "math is the only valid tool"
- "math is always and everywhere self-evidently applicable"
- "math is all that really counts"
- "if it's not in math, it's not really economics'
- "almost everything can be adequately understood and analyzed with math"

Mainstream economists have always wanted to use their hammer, and so have decided to pretend that the world looks like a nail. Pretending that uncertainty can be reduced to risk and that all activities, relations, processes, and events can be adequately converted to pure numbers, have only contributed to making economics irrelevant and powerless when confronting real-world financial crises and economic havoc.

Mainstream economic theory today is still in the storytelling business whereby economic theorists create mathematical make-believe analog models of the target system -- usually conceived as the real

economic system. This mathematical modeling activity is considered useful and essential. To understand and explain relations between different entities in the real economy the predominant strategy is to build mathematical models and make things happen in these 'analog-economy models' rather than engineering things happening in real economies.

Without strong evidence, all kinds of absurd claims and nonsense may pretend to be science. But -- math never has, and never will, be able to establish the truth value of facts.

We have to demand more of a justification than rather watered-down versions of 'anything goes' when it comes to the main postulates on which mainstream economics is founded. If one proposes 'efficient markets' or 'rational expectations' one also has to support their underlying assumptions. As a rule, none is given, which makes it rather puzzling how things like 'efficient markets' and 'rational expectations' have become standard modeling assumptions made in much of modern macroeconomics. The reason for this sad state of 'modern' economics is that economists often mistake mathematical beauty for truth.

How do we put an end to this intellectual cataclysm? How do we re-establish credence and trust in economics as a science? A couple of changes are absolutely decisive:

- Stop pretending that we have exact and rigorous answers on everything. Because we don't. We build models and theories and tell people that we can calculate and foresee the future. But we do this based on mathematical and statistical assumptions that often have little or nothing to do with reality. By pretending that there is no really important difference between model and reality we lull people into thinking that we have things under control. We haven't! This false feeling of security was one of the factors that contributed to the financial crisis of 2008.

- Stop the childish and exaggerated belief in mathematics giving answers to important economic questions. Mathematics gives exact answers to exact questions. But the relevant and interesting questions we face in the economic realm are rarely of that kind. Questions like "Is 2 + 2 = 4?' are never posed in real economies. Instead of a fundamentally misplaced reliance on abstract mathematical-deductive-axiomatic models having anything of substance to contribute to our knowledge of real economies, it would be far better if we pursued "thicker' models and relevant empirical studies and observations.

- Stop pretending that there are laws in economics. There are no universal laws in economics. Economies are not like planetary systems or physics labs. The most we can aspire to in real economies is establishing possible tendencies with varying degrees of generalizability.

- Stop treating other social sciences as poor relations. Economics has long suffered from hubris. A more broad-minded and multifarious science would enrich today's altogether too autistic economics.

- Stop building models and making forecasts of the future based on totally unreal micro-founded macro models with intertemporally optimizing robot-like representative actors equipped with rational expectations. This is pure nonsense. We have to build our models on assumptions that are not so blatantly in contradiction to reality. Assuming that people are green and come from Mars is not a good – not even as a 'successive approximation' – modeling strategy.

Model and Reality

Mainstream economists often believe that models and theories will approach the 'truth' through 'successive approximations' by incorporating more and more factors and relaxing the idealizing assumptions. However, using 'isolation' and 'successive approximations' as a heuristic method is not convincing. For the method to work well, the factors studied in 'isolation' must be real causal factors or tendencies, and the effects of the studied factors must be treated separately and mechanically added and interacted. However, rationality and maximization assumptions in mainstream theory do not represent real causal relationships or tendencies, but often only misleading fictions. It is difficult to imagine that complex economic phenomena could be adequately treated by considering them as if they existed separately or in isolation from each other. Economic contexts are fundamentally contextual and open. Therefore, the possibility of treating the economy as if its causal factors could be mechanically combined is generally excluded. It is not clear what kind of interesting and relevant conclusions can be drawn about our economy using 'successive approximations' from assumptions about all-knowing and infallible individuals and companies. If the starting point is wrong, so is the end product. The analysis does not contribute to explaining the generative mechanisms and forces of society and the economy. It only leads us astray.

Economics is more model-oriented than any other social science. There are many reasons for this - the history of the subject, ideals (drawn from the natural sciences), universal claims, the desire to explain as much as possible with as little as possible, rigor, precision, etc.

The approach is fundamentally *analytical* -- the whole is broken down into its components so that it is possible to explain (reduce) the aggregate (macro) as a result of the interaction between the parts (micro).

Mainstream economists typically base their models on a number of core assumptions (CA) -- which fundamentally describe actors as 'rational' -- as well as a number of auxiliary assumptions (AA). Together, (CA) and (AA) constitute what we could call the base model (M) for all mainstream models. Based on these two sets of assumptions, one tries to explain and predict both individual (micro) and societal phenomena (macro).

The core assumptions typically consist of:

CA_1 Completeness -- the rational actor is always able to compare different alternatives and determine which she prefers
CA_2 Transitivity -- if the actor prefers A to B and B to C, she must prefer A to C
CA_3 Non-satiation -- more is always better than less
CA_4 Maximization of expected utility -- in situations characterized by risk, the actor always maximizes expected utility
CA_5 Consistent economic equilibria -- different actors' actions are consistent, and their interaction results in an equilibrium.

When describing actors as rational in these models, *instrumental* rationality is meant -- the actors are assumed to choose alternatives that have the best consequences given their given preferences. How these given preferences have arisen is generally perceived to be outside the scope of the concept of rationality and therefore not part of economic theory as such.

The picture one gets of the core assumptions ('rational choices') is a rational actor with strong cognitive capacities, who knows what she wants, carefully considers her alternatives, and given her preferences,

chooses what she believes has the best consequences for her. By weighing the various alternatives against each other, the actor makes a consistent, rational choice and acts accordingly.

The auxiliary assumptions (AA) specify spatial and temporal aspects of the type of interaction that can take place between 'rational' actors. These assumptions often provide answers to questions such as:

AA_1: Who are the actors, and where and when do they interact?
AA_2: What are their goals and aspirations?
AA_3: What interests do they have?
AA_4: What are their expectations?
AA_5: What kind of agency do they have?
AA_6: What kind of agreements can they make?
AA_7: How much and what kind of information do they possess?
AA_8: How do their actions interact with each other?

The basic model for all mainstream models is thus a general picture of what (axiomatically) constitutes optimizing rational actors (CA), and a more specific description (AA) of the situations in which these actors act (which means that AA functions as a restriction that determines the intended application domain for CA and the deductively derived theorems). The list of assumptions can never be complete because there are always unspecified 'background assumptions' and unmentioned omissions, often based on some kind of negligibility and application considerations. The hope is that this 'thin' set of assumptions will be sufficient to explain and predict 'rich' phenomena in the real, complex world.

In the extreme case, the theorems turn into non-testable tautological thought experiments with no other empirical ambitions than to tell a coherent fictional 'as-if' history.

Not clearly distinguishing between (CA) and (AA) opens up all sorts of attempts to 'save' or 'immunize' models from criticism by unfairly 'sliding' between interpreting the models as empirically empty deductive-axiomatic analytical 'systems' or as models with explicit empirical aspirations (cf. Albert 2012). In ordinary cases, flexibility may be seen as positive, but in a methodological context, it is rather a sign of a problem. Models that are compatible with everything or come with unspecified application domains are worthless from a scientific point of view.

Economics -- unlike logic and mathematics -- should be an empirical science, and empirical tests of 'axioms' should obviously be relevant to such a discipline. Even if mainstream economists themselves (implicitly or explicitly) claim that their axioms are universally accepted as 'true' and without the need for proof, this is obviously not a reason for others to simply accept them.

When mainstream economists' deductive 'thinking' is used, it usually results in the construction of 'as-if' models based on some form of idealization logic and a set of axiomatic assumptions from which consistent and precise inferences can be made. The beauty of this, of course, is that if the axiomatic premises are true, the conclusions necessarily follow. However, although the procedure is successfully used in mathematics and mathematically derived axiomatic-deductive systems, it is a poor guide for understanding and explaining systems in the real world.

Most theoretical models that mainstream economists work with are abstract and unrealistic constructions that are used to construct non-testable hypotheses. How this can tell us anything relevant and interesting about the world we live in is difficult to see. Faced with the massive empirical failures these models and theories have led to, many mainstream economists retreat and choose to present their models and theories as merely a kind of thought experiment without any real aspirations to tell us

anything about the real world. Instead of 'bridging' the model and reality, they simply give up. However, this type of scientific defeatism is completely unacceptable. It can never be enough to prove or deduce things in a model world. If theories do not -- directly or indirectly -- tell us something about the world we live in, why should we waste time on them?

Open and closed models

In all science, the search for regularities and laws has been an essential feature. The goal has often been to make predictions based on these regularities and laws. Although natural science has been partly successful in this regard, economics has encountered major problems (which perhaps testify to all the failed economic forecasts). Some have drawn the conclusion that economic science is still underdeveloped, and if given enough time, it should certainly be possible to achieve the same predictive reliability as the natural sciences. From a critical realist perspective, however, the economic science's lack of predictive ability is mainly a result of its research object's inherent structure and character.

The critical realist researcher investigates the conditions that must be met for regularities and law-like regularities to be said to exist (cf. Lawson 1997). One condition is that there must be no changes or qualitative variations in the object that determines the causal relationships, i.e., the underlying mechanisms must be stable (the condition of *internal closure*). The internal structure of the system must be constant over time. One must try to find an analysis unit that is invariant over time and that exhibits a constant pattern of behavior. This makes it possible to postulate a stable and identifiable relationship between a set of conditions X_1, X_2, ..., X_n and the value of a variable Y.

Another condition for regularities and law-like regularities is that the relationship between the causal mechanisms and the mechanisms in the environment that influence them must be constant (the condition of *external closure*). If we are to identify empirical regularities, we must be able to isolate the system from non-constant external influences. This means that only explicitly treated conditions X_1, X_2, ..., X_n have a systematic, non-constant impact on Y. The purpose of this condition is to exclude the impact of unspecified conditions on Y. If a system cannot be physically isolated from the environment, the environmental factors must either be internalized in the system, or one must ensure that they have a constant and unchanging influence on the system. If both closure conditions are met, we have a closed system that meets the conditions for regularities to be produced. But most real systems we encounter do not meet these conditions. They are *open* systems whose possible regularities are more transient, local, or approximate.

The pitfalls of econometrics

As an example of the problems that the deductivist approach leads to, econometrics is a good example. This part of economics can be said to aim to establish probabilistic regularities consisting of a dependent variable Y being functionally related to a set of independent variables X, so that changes in the latter can be shown to give rise to predictable variations in the former. To date, no universal laws have been found - judging from the existing literature. Today, we are still as surprised as Trygve Haavelmo was in the 1940s that the estimated relationships break down at the same rate as new data becomes available.

From a critical realist perspective, however, it is hardly surprising that an approach that assumes a closed system fails when applied to an open system such as an economy. According to the so-called 'Lucas critique', the problem with econometric models is that the identified relationships are unstable,

while the underlying theory is based on the assumed existence of stable parameter relationships. Therefore, the models do not allow for more far-reaching predictions.

The starting point for econometric analysis is partly that social and economic relations can be modeled using mathematical functions, and partly that it is possible to obtain knowledge about the parameters of these functions. However, it is difficult to confirm the validity of these assumptions. Such confirmation would imply that the econometrician not only had a statistical model but also a 'true model', a mathematical function that connects the different variables. Therefore, researchers often assume that the true model exists but is unknown. Instead of trying to estimate how well the econometric model coincides with the 'true' model, researchers are content to analyze how well the econometric model specifies a particular theory. It is difficult to see how this could contribute in any interesting way to explaining real events and structures.

The pre-requisite for predictions based on econometric equations to have any greater value is, of course, that the relationships represented in the equations are stable over time and space. If the mechanisms through which the relationships operate do not remain constant over time, reliable predictions cannot be made. In the economy, it is a pervasive characteristic that behavioral patterns do not exhibit the required invariance. The causal forces and structures that shape the relationships are not constant but constantly evolving and transforming. Consumption and investment patterns fluctuate continuously, and this is not solely due to relevant variables being omitted or mis-specified in the econometric model. The relationships themselves change due to variations in the underlying mechanisms. The error in this type of generalization lies simply in treating temporary and transient relationships as universal and invariant natural laws.

The very basic assumption for applying probabilistic logic to real systems is that the uncertainty exhibited is not of the type that we usually call *genuine uncertainty* (where we cannot practically assign any numerical values to future expectations). Instead, all uncertainty is assumed to be reducible to *calculable risk* (such as coin tossing and roulette). That this would be a sustainable assumption for understanding real economic systems and relationships is simply a belief for which there is no other foundation than the belief itself.

Trying to reduce the risk of having established only 'spurious relations' when dealing with observational data, statisticians and econometricians standardly add control variables. The hope is that one thereby will be able to make more reliable causal inferences. But — as Keynes showed already back in the 1930s when criticizing statistical-econometric applications of regression analysis — if you do not manage to get hold of *all* potential confounding factors, the model risks producing estimates of the variable of interest that are even worse than models without any control variables at all.

Conclusion: think twice before you simply include 'control variables' in your models!

'Kitchen sink' econometric models are often the result of researchers trying to control for confounding. But what they usually haven't understood is that the confounder problem requires a causal solution and not statistical 'control.' Controlling for everything opens up the risk that we control for 'collider' variables and thereby create 'back-door paths' which gives us confounding that wasn't there to begin with.

Causality can never be reduced to a question of statistics or probabilities unless you are — miraculously — able to keep constant all other factors that influence the probability of the outcome studied. To understand causality, we always have to relate it to a specific causal structure. Statistical correlations are never enough. No structure, no causality.

Explanatory fictionalism and the experimental hype

One of the limitations of economics is the restricted possibility to perform experiments, forcing it to mainly rely on observational studies for knowledge of real-world economies.

But still — the idea of performing laboratory experiments holds a firm grip on our wish to discover (causal) relationships between economic 'variables.' If we only could isolate and manipulate variables in controlled environments, we would probably find ourselves in a situation where we with greater 'rigour' and 'precision' could describe, predict, or explain economic happenings in terms of 'structural' causes, 'parameter' values of relevant variables, and economic 'laws.'

Galileo Galilei's experiments are often held as exemplary for how to perform experiments to learn something about the real world. Galileo's heavy balls dropping from the tower of Pisa, confirmed that the distance an object falls is proportional to the square of time and that this law (empirical regularity) of falling bodies could be applicable outside a vacuum tube when e. g. air existence is negligible.

The big problem is to decide or find out exactly for which objects air resistance (and other potentially 'confounding' factors) is 'negligible.' In the case of heavy balls, air resistance is obviously negligible, but how about feathers or plastic bags?

One possibility is to take the all-encompassing-theory road and find out all about possible disturbing/confounding factors — not only air resistance — influencing the fall and build that into one great model delivering accurate predictions on what happens when the object that falls is not only a heavy ball but feathers and plastic bags. This usually amounts to ultimately stating some kind of ceteris paribus interpretation of the 'law.'

Another road to take would be to concentrate on the negligibility assumption and to specify the domain of applicability to be only heavy compact bodies. The price you have to pay for this is that (1) 'negligibility' may be hard to establish in open real-world systems, (2) the generalization you can make from 'sample' to 'population' is heavily restricted, and (3) you actually have to use some 'shoe leather' and empirically try to find out how large is the 'reach' of the 'law.'

In mainstream economics, one has usually settled for the 'theoretical' road (and in case you think the present 'natural experiments' hype has changed anything, remember that to mimic real experiments, exceedingly stringent special conditions standardly have to obtain).

In the end, it all boils down to one question — are there any Galilean 'heavy balls' to be found in economics, so that we can indisputably establish the existence of economic laws operating in real-world economies?

As far as I can see there are some heavy balls out there, but not even one single real economic law.

Economic factors/variables are more like feathers than heavy balls — non-negligible factors (like air resistance and chaotic turbulence) are hard to rule out as having no influence on the object studied.

Galilean experiments are hard to carry out in economics, and the theoretical 'analogue' models economists construct and in which they perform their 'thought experiments' build on assumptions that are far away from the kind of idealized conditions under which Galileo performed his experiments. The 'nomological machines' that Galileo and other scientists have been able to construct have no real analogues in economics. The stability, autonomy, modularity, and interventional invariance, that we

may find between entities in nature, simply are not there in real-world economies. That's a real-world fact, and contrary to the beliefs of most mainstream economists, they won't go away simply by applying deductive-axiomatic economic theory with more or less unsubstantiated assumptions.

By this, I do not mean to say that we have to discard all (causal) theories/laws building on modularity, stability, invariance, etc. But we have to acknowledge the fact that outside the systems that possibly fulfil these requirements/assumptions, they are of little substantial value. Running paper and pen experiments on artificial 'analogue' model economies is a sure way of 'establishing' (causal) economic laws or solving intricate econometric problems of autonomy, identification, invariance and structural stability — in the model world. But they are pure substitutes for the real thing and they don't have much bearing on what goes on in real-world open social systems. Setting up convenient circumstances for conducting Galilean experiments may tell us a lot about what happens under those kinds of circumstances. But — few, if any, real-world social systems are 'convenient.' So, most of those theories and models, are irrelevant for letting us know what we really want to know.

To solve, understand, or explain real-world problems you actually have to know something about them — logic, pure mathematics, data simulations or deductive axiomatics don't take you very far. Most econometrics and economic theories/models are splendid logic machines. But — applying them to the real world is a totally hopeless undertaking! The assumptions one has to make in order to successfully apply these deductive-axiomatic theories/models/machines are devastatingly restrictive and mostly empirically untestable– and hence make their real-world scope ridiculously narrow. To fruitfully analyze real-world phenomena with models and theories you cannot build on patently and known to be ridiculously absurd assumptions. No matter how much you would like the world to entirely consist of heavy balls, the world is not like that. The world also has its fair share of feathers and plastic bags.

Most of the 'idealizations' we find in mainstream economic models are not 'core' assumptions, but rather structural 'auxiliary' assumptions. Without those supplementary assumptions, the core assumptions deliver next to nothing of interest. So, to come up with interesting conclusions you have to rely heavily on those other — 'structural' — assumptions.

In physics, we have theories and centuries of experience and experiments that show how gravity makes bodies move. In economics, we know there is nothing equivalent. So instead, mainstream economists necessarily have to load their theories and models with sets of auxiliary structural assumptions to get any results at all in their models.

So why then do mainstream economists keep on pursuing this modelling project?

The way axioms and theorems are formulated in mainstream economics often leaves their specification without almost any restrictions whatsoever, safely making every imaginable evidence compatible with the all-embracing 'theory' — and theory without informational content never risks being empirically tested and found falsified. Used in mainstream 'thought experimental' activities, it may, of course, be very 'handy,' but totally void of any empirical value.

Some economic methodologists have lately been arguing that economic models may well be considered 'minimal models' that portray 'credible worlds' without having to care about things like similarity, isomorphism, simplified 'representationality' or resemblance to the real world. These models are said to resemble 'realistic novels' that portray 'possible worlds'. And sure: economists constructing and working with those kinds of models learn things about what might happen in those 'possible worlds'. But is that really the stuff real science is made of? I think not. As long as one doesn't come up with credible export warrants to real-world target systems and show how those models — often building on

idealizations with known to be false assumptions — enhance our understanding or explanations about the real world, well, they are just nothing more than just novels. Showing that something is possible in a 'possible world' doesn't give us a justified license to infer that it therefore also is possible in the real world. 'The Great Gatsby' is a wonderful novel, but if you truly want to learn about what is going on in the world of finance, I recommend rather reading Minsky or Keynes and directly confronting real-world finance.

Different models have different cognitive goals. Constructing models that aim for explanatory insights may not optimize the models for making (quantitative) predictions or deliver some kind of 'understanding' of what's going on in the intended target system. All modelling in science has tradeoffs. There simply is no 'best' model. For one purpose in one context model A is 'best', for other purposes and contexts model B may be deemed 'best'. Depending on the level of generality, abstraction, and depth, we come up with different models. But even so, one could argue that if we are looking for 'adequate' explanations, it is not enough to just come up with 'minimal' or 'credible world' models.

The assumptions and descriptions we use in our modelling have to be true — or at least 'harmlessly' false — and give a sufficiently detailed characterization of the mechanisms and forces at work. Models in mainstream economics do nothing of the kind.

Coming up with models that show how things may possibly be explained is not what we are looking for. It is not enough. We want to have models that build on assumptions that are not in conflict with known facts and that show how things actually are to be explained. Our aspirations have to be more far-reaching than just constructing coherent and 'credible' models about 'possible worlds'. We want to understand and explain 'difference-making' in the real world and not just in some made-up fantasy world. No matter how many mechanisms or coherent relations you represent in your model, you still have to show that these mechanisms and relations are at work and exist in society if we are to do real science. Science has to be something more than just more or less realistic 'storytelling' or 'explanatory fictionalism.' You have to provide decisive empirical evidence that what you can infer in your model also helps us to uncover what actually goes on in the real world. It is not enough to present epistemically informative insights about logically possible models. You also, and more importantly, have to have a world-linking argumentation and show how those models explain or teach us something about real-world economies. If you fail to support your models in that way, why should we care about them? And if you do not inform us about what are the real-world intended target systems of your modelling, how are we going to be able to value or test them? Without giving that kind of information it is impossible for us to check if the 'possible world' models you come up with actually hold also for the one world in which we live — the real world.

Keynes' critique of econometrics

Mainstream economists often hold the view that Keynes' criticism of econometrics was the result of a sadly misinformed and misguided person who disliked and did not understand much of it. This is, however, nothing but a gross misapprehension (cf. Nasir & Morgan 2023). To be careful and cautious is not the same as to dislike. Keynes did not misunderstand the crucial issues at stake in the development of econometrics. Quite the contrary. He knew them all too well — and was not satisfied with the validity and philosophical underpinning of the assumptions made for applying its methods.

Keynes' critique of the 'logical issues' regarding the conditions that have to be satisfied if we are going to be able to apply econometric methods, is still valid and unanswered in the sense that the problems

he pointed at are still with us today and largely unsolved. Ignoring them — the most common practice among applied econometricians — is not to solve them.

To apply statistical and mathematical methods to the real-world economy, the econometrician has to make some quite strong assumptions. In a review of Tinbergen's econometric work, Keynes (1939) gave a comprehensive critique of Tinbergen's work, focusing on the limiting and unreal character of the assumptions that econometric analyses build on:

- Completeness: Where Tinbergen attempts to specify and quantify which factors influence the business cycle, Keynes maintains there must be a complete list of all the relevant factors to avoid misspecification and spurious causal claims. Usually, this problem is 'solved' by econometricians assuming that they somehow have a 'correct' model specification. Keynes was, to put it mildly, unconvinced.

- Homogeneity: To make inductive inferences possible — and be able to apply econometrics — the system we try to analyse has to have a large degree of 'homogeneity.' According to Keynes most social and economic systems — especially from the perspective of real historical time — lack that 'homogeneity.' As he had argued already in Treatise on Probability, it wasn't always possible to take repeated samples from a fixed population when we were analyzing real-world economies. In many cases, there simply are no reasons at all to assume the samples to be homogenous. Lack of 'homogeneity' makes the principle of 'limited independent variety' non-applicable, and hence makes inductive inferences, strictly seen, impossible since one of its fundamental logical premises is not satisfied.

And then, of course, there is also the 'reverse' variability problem of non-excitation: factors that do not change significantly during the period analysed, can still very well be extremely important causal factors.

- Stability: Tinbergen assumes there is a stable spatio-temporal relationship between the variables his econometric models analyze. But as Keynes had argued already in his Treatise on Probability it was not really possible to make inductive generalisations based on correlations in one sample. As later studies of 'regime shifts' and 'structural breaks' have shown us, it is exceedingly difficult to find and establish the existence of stable econometric parameters for anything but rather short time series.

- Measurability: Tinbergen's model assumes that all relevant factors are measurable. Keynes questions if it is possible to adequately quantify and measure things like expectations and political and psychological factors. And more than anything, he questioned — both on epistemological and ontological grounds — that it was always and everywhere possible to measure real-world uncertainty with the help of probabilistic risk measures. Thinking otherwise can, as Keynes wrote, "only lead to error and delusion."

- Independence: Tinbergen assumes that the variables he treats are independent (still a standard assumption in econometrics). Keynes argues that in such a complex, organic and evolutionary system as an economy, independence is a deeply unrealistic assumption to make. Building econometric models from that kind of simplistic and unrealistic assumptions risks producing nothing but spurious correlations and causalities. Real-world economies are organic systems for which the statistical methods used in econometrics are ill-suited, or even, strictly seen, inapplicable. Mechanical probabilistic models have little leverage when applied to non-atomic evolving organic systems — such as economies.

Building econometric models can't be a goal in itself. Good econometric models are means that make it possible for us to infer things about the real-world systems they 'represent.' If we can't show that the mechanisms or causes that we isolate and handle in our econometric models are 'exportable' to the real world, they are of limited value to our understanding, explanations or predictions of real-world economic systems.

- Linearity: To make his models tractable, Tinbergen assumes the relationships between the variables he studies to be linear. This is still standard procedure today, but to Keynes, it was a 'fallacy of reification' to assume that all quantities are additive (an assumption closely linked to independence and linearity).

And as even one of the founding fathers of modern econometrics — Trygve Haavelmo — wrote (Haavelmo, 1944, 6): "What is the use of testing, say, the significance of regression coefficients, when maybe, the whole assumption of the linear regression equation is wrong?"

Real-world social systems are usually not governed by stable causal mechanisms or capacities. The kinds of 'laws' and relations that econometrics has established, are laws and relations about entities in models that presuppose causal mechanisms and variables — and the relationship between them — being linear, additive, homogenous, stable, invariant and atomistic. But — when causal mechanisms operate in the real world they only do it in ever-changing and unstable combinations where the whole is more than a mechanical sum of parts. Since statisticians and econometricians — as far as I can see — haven't been able to convincingly warrant their assumptions of homogeneity, stability, invariance, independence, and additivity as being ontologically isomorphic to real-world economic systems, Keynes' critique is still valid. There are strong reasons to remain doubtful of the scientific aspirations of econometrics. Especially when it comes to using econometrics for making causal inferences, it is still often based on counterfactual assumptions that have outrageously weak grounds.

In his critique of Tinbergen, Keynes points us to the fundamental logical, epistemological and ontological problems of applying statistical methods to a basically unpredictable, uncertain, complex, unstable, interdependent, and ever-changing social reality. Methods designed to analyse repeated sampling in controlled experiments under fixed conditions are not easily extended to an organic and non-atomistic world where time and history play decisive roles.

Econometric modelling should never be a substitute for thinking. From that perspective, it is really depressing to see how much of Keynes' critique of the pioneering econometrics in the 1930s-1940s is still relevant today.

Randomization tools

Since a couple of decades, we have seen a new trend in economics, where there is a growing interest in experiments and — not least — how to design them to possibly provide answers to questions about causality and policy effects. Economic research on discrimination nowadays often emphasizes the importance of a randomization design, for example when trying to determine to what extent discrimination can be causally attributed to differences in preferences or information, using so-called correspondence tests and field experiments.

A common starting point is the 'counterfactual approach' developed mainly by Neyman and Rubin, which is often presented and discussed based on examples of randomized control studies, natural experiments, difference in difference, matching, regression discontinuity, etc. Mainstream economists

generally view this development of the economics toolbox positively. However, I — like, for example, Nancy Cartwright and Angus Deaton — am not entirely positive about the randomization approach.

A notable limitation of counterfactual randomization designs is that they only give us answers on how 'treatment groups' differ on average from 'control groups.' Let me give an example to illustrate how limiting this fact can be:

> Among school debaters and politicians, it is often claimed that so-called 'independent schools' (charter schools) are better than municipal schools. They are said to lead to better results. To find out if this is really the case, a number of students are randomly selected to take a test. The result could be: Test result = 20 + 5T, where T=1 if the student attends an independent school and T=0 if the student attends a municipal school. This would confirm the assumption that independent school students have an average of 5 points higher results than students in municipal schools. Now, politicians (hopefully) are aware that this statistical result cannot be interpreted in causal terms because independent school students typically do not have the same background (socio-economic, educational, cultural, etc.) as those who attend municipal schools (the relationship between school type and result is confounded by selection bias). To obtain a better measure of the causal effects of school type, politicians suggest that 1000 students be admitted to an independent school through a lottery — a classic example of a randomization design in natural experiments. The chance of winning is 10%, so 100 students are given this opportunity. Of these, 20 accept the offer to attend an independent school. Of the 900 lottery participants who do not 'win,' 100 choose to attend an independent school. The lottery is often perceived by school researchers as an 'instrumental variable,' and when the analysis is carried out, the result is: Test result = 20 + 2T. This is standardly interpreted as having obtained a causal measure of how much better students would, on average, perform on the test if they chose to attend independent schools instead of municipal schools. But is it true? No! If not all school students have exactly the same test results (which is a rather far-fetched 'homogeneity assumption'), the specified average causal effect only applies to the students who choose to attend an independent school if they 'win' the lottery, but who would not otherwise choose to attend an independent school (in statistical jargon, we call these 'compliers'). It is difficult to see why this group of students would be particularly interesting in this example, given that the average causal effect estimated using the instrumental variable says nothing at all about the effect on the majority (the 100 out of 120 who choose to attend an independent school without 'winning' in the lottery) of those who choose to attend an independent school.

Conclusion: Researchers must be much more careful in interpreting 'average estimates' as causal. Reality exhibits a high degree of heterogeneity, and 'average parameters' often tell us very little!

To randomize ideally means that we achieve orthogonality (independence) in our models. But it does not mean that in real experiments when we randomize, we achieve this ideal. The 'balance' that randomization should ideally result in cannot be taken for granted when the ideal is translated into reality. Here, one must argue and verify that the 'assignment mechanism' is truly stochastic and that 'balance' has indeed been achieved!

Even if we accept the limitation of only being able to say something about average treatment effects there is another theoretical problem. An ideal randomized experiment assumes that a number of individuals are first chosen from a randomly selected population and then randomly assigned to a

treatment group or a control group. Given that both selection and assignment are successfully carried out randomly, it can be shown that the expected outcome difference between the two groups is the average causal effect in the population. The hitch is that the experiments conducted almost never involve participants selected from a random population! In most cases, experiments are started because there is a problem of some kind in a given population (e.g., schoolchildren or job seekers in country X) that one wants to address. An ideal randomized experiment assumes that *both* selection and assignment are randomized — this means that virtually none of the empirical results that randomization advocates so eagerly tout hold up in a strict mathematical-statistical sense. The fact that only assignment is talked about when it comes to 'as if' randomization in natural experiments is hardly a coincidence. Moreover, when it comes to 'as if' randomization in natural experiments, the sad but inevitable fact is that there can always be a dependency between the variables being studied and unobservable factors in the error term, which can never be tested!

Another significant and major problem is that researchers who use these randomization-based research strategies often, in order to achieve 'exact' and 'precise' results, set up problem formulations that are not at all the ones we really want answers to. Design becomes the main thing, and as long as one can get more or less clever experiments in place, they believe they can draw far-reaching conclusions about both causality and the ability to generalize experimental outcomes to larger populations. Unfortunately, this often means that this type of research has a negative bias away from interesting and important problems towards prioritizing method selection. Design and research planning are important, but the credibility of research ultimately lies in being able to provide answers to relevant questions that both citizens and researchers want answers to.

Believing there is only one really good evidence-based method on the market — and that randomization is the only way to achieve scientific validity — blinds people to searching for and using other methods that in many contexts are better. Insisting on using only one tool often means using the wrong tool.

The 'true model' assumption

Most work in econometrics is — still — made on the assumption that the researcher has a theoretical model that is 'true.' Based on this belief of having a correct specification for an econometric model or running a regression, one proceeds as if the only problem remaining to solve has to do with measurement and observation.

When things sound too good to be true, they usually aren't. And that goes for econometric wet dreams too. The snag is that there is pretty little to support the perfect specification assumption. Looking around in social science and economics we don't find a single regression or econometric model that lives up to the standards set by the 'true' theoretical model — and there is almost nothing that gives us reason to believe things will be different in the future.

To think that we are being able to construct a model where all relevant variables are included and correctly specify the functional relationships that exist between them is not only a belief without support but a belief impossible to support.

The theories we work with when building our econometric regression models are insufficient. No matter what we study, there are always some variables missing, and we don't know the correct way to functionally specify the relationships between the variables.

Every regression model constructed is mis-specified. There is always an endless list of possible variables to include, and endless possible ways to specify the relationships between them. So every applied econometrician comes up with his own specification and 'parameter' estimates. The econometric Holy Grail of consistent and stable parameter values is nothing but a dream.

A rigorous application of econometric methods in economics really presupposes that the phenomena of our real-world economies are ruled by stable causal relations between variables. Parameter values estimated in specific spatio-temporal contexts are presupposed to be exportable to different contexts. To warrant this assumption one, however, has to convincingly establish that the targeted acting causes are stable and invariant so that they maintain their parametric status after the bridging. The endemic lack of predictive success of the econometric project indicates that this hope of finding fixed parameters is a hope for which there really is no other ground than hope itself.

The theoretical conditions that must be fulfilled for regression analysis and econometrics to really work are nowhere even closely met in reality. Making outlandish statistical assumptions does not provide a solid ground for doing relevant social science and economics. Although regression analysis and econometrics have become the most used quantitative methods in social sciences and economics today, it's still a fact that almost all of the inferences made from them are invalid.

Given the usual set of assumptions (such as manipulability, transitivity, separability, additivity, linearity, etc.) econometrics delivers deductive inferences. But the problem, of course, is that we will never completely know when the assumptions are right. Conclusions can only be as certain as their premises — and that also applies to econometrics.

Although 'ideally controlled experiments' may tell us with certainty what causes what effects, this is so only when given the right 'closures.' Making appropriate extrapolations from (ideal, accidental, natural or quasi) experiments to different settings, populations or target systems, is not easy. "It works there' is no evidence for "it will work here.' The causal background assumptions made must be justified, and without licenses to export, the value of 'rigorous' and 'precise' methods used when analyzing 'natural experiments' is often despairingly small. Just to take one example — since the core assumptions on which instrumental variables analysis builds are *never* directly testable, those of us who choose to use instrumental variables to find out about causality *always* have to defend and argue for the validity of the assumptions the causal inferences build on. Especially when dealing with natural experiments, we should be very cautious when being presented with causal conclusions without convincing arguments about the veracity of the assumptions made. If you are out to make causal inferences, you have to rely on a trustworthy theory of the data-generating process. The empirical results causal analysis supplies us with are only as good as the assumptions we make about the data-generating process. Garbage in, garbage out.

Non-manipulability and the limits of potential outcome models

Framing all causal questions as questions of manipulation or intervention runs into many problems, especially when we open up for 'hypothetical' and 'symbolic' interventions. Humans have few barriers to imagining things, but that often also makes it difficult to value the proposed thought experiments in terms of relevance. Performing 'well-defined' interventions is one thing, but if we do not want to give up searching for answers to the questions we are interested in and instead only search for answerable questions, interventionist studies are of limited applicability and value. Intervention effects in thought experiments are not self-evidently the causal effects we are looking for. Identifying causes (reverse causality) and measuring effects of causes (forward causality) is not the same. In social sciences, like

economics, we standardly first try to identify the problem and why it occurred, and then afterwards look at the effects of the causes.

Leaning on the interventionist approach often means that instead of posing interesting questions on a social level, focus is on individuals (cf. Goldthorpe 2001). Instead of asking about structural socio-economic factors behind, e.g., gender or racial discrimination, the focus is on the choices individuals make (which also tends to make the explanations presented inadequately 'deep'). Within the manipulation approach to causality you are only allowed to ask certain rather restricted causal questions. A typical example of the dangers of this limiting approach is 'Nobel prize' winner Esther Duflo, who thinks that economics should be based on evidence from randomised experiments and field studies. Duflo *et consortes* want to give up on 'big ideas' like political economy and institutional reform and instead go for solving more manageable problems the way plumbers do (cf. Duflo 2017). Yours truly is far from sure that is the right way to move economics forward and make it a relevant and realist science. A plumber can fix minor leaks in your system, but if the whole system is rotten, something more than good old fashion plumbing is needed. The big social and economic problems we face today are not going to be solved by plumbers performing interventions or manipulations in the form of randomised control trials.

Although, of course, it is possible (more or less, depending on context) to retrofit causal questions into a manipulation/intervention framework, before we are there, we have to agree on having identified the causal problem we try to deal with when recommending different policies. Before we can calculate causal effects we have to identify the causes, and this is perhaps not always best done within a manipulation/intervention framework. One problem is that the manipulation/intervention approach in broader social and economic contexts requires a reframing of the questions we pose, which often means that we get 'well-defined' causal answers, but not necessarily answers to the questions we really are interested in finding answers to. The manipulation/intervention framework is one way to do causal analysis. But it is not the way to do it. Being an advocate of 'inference to the best explanation' I think we also have to more carefully consider explanatory considerations when estimating and identifying causal relations.

A popular idea in quantitative social sciences is to think of a cause (C) as something that increases the probability of its effect or outcome (O). That is:

$$P(O|C) > P(O|-C).$$

However, as is also well-known, a correlation between two variables, say A and B, does not necessarily imply that that one is a cause of the other, or the other way around, since they may both be an effect of a common cause, C.

In statistics and econometrics, we usually solve this "confounder' problem by 'controlling for' C, i. e. by holding C fixed. This means that we actually look at different "populations' – those in which C occurs in every case, and those in which C doesn't occur at all. This means that knowing the value of A does not influence the probability of C [P(C|A) = P(C)]. So, if there then still exists a correlation between A and B in either of these populations, there has to be some other cause operating. But if all other possible causes have been 'controlled for' too, and there is still a correlation between A and B, we may safely conclude that A is a cause of B, since by 'controlling for' all other possible causes, the correlation between the putative cause A and all the other possible causes (D, E,. F …) is broken.

This is of course a very demanding prerequisite, since we may never actually be sure to have identified all putative causes. Even in scientific experiments, the number of uncontrolled causes may be

innumerable. Since nothing less will do, we do all understand how hard it is to actually get from correlation to causality. This also means that only relying on statistics or econometrics is not enough to deduce causes from correlations.

Some people think that randomization may solve the empirical problem. By randomizing we are getting different 'populations' that are homogeneous in regards to all variables except the one we think is a genuine cause. In that way, we are supposed to be able not to actually have to know what all these other factors are.

If you succeed in performing an ideal randomization with different treatment groups and control groups that is attainable. But it presupposes that you really have been able to establish – and not just assume – that the probability of all other causes but the putative (A) have the same probability distribution in the treatment and control groups, and that the probability of assignment to treatment or control groups are independent of all other possible causal variables.

Unfortunately, *real* experiments and *real* randomizations seldom or never achieve this. So, yes, we may do without knowing all causes, but it takes ideal experiments and ideal randomizations to do that, not real ones. That means that in practice we do have to have sufficient background knowledge to deduce causal knowledge. Without old knowledge, we can't get new knowledge. No causes in, no causes out.

Once you include all actual causes into the original (over)simplified model, it may well be that the causes are no longer independent or linear and that *a fortiori* the coefficients in the econometric equations no longer are identifiable. And so, since all causal factors are not included in the original econometric model, it is not an adequate representation of the real causal structure of the economy that the model is purportedly meant to represent.

Econometrics is basically a deductive method. Given the assumptions (such as manipulability, transitivity, exchangeability, monotonicity, ignorability, Reichenbach probability principles, separability, additivity, linearity etc) it delivers deductive inferences. The problem, of course, is that we will never completely know when the assumptions are right. Real target systems are seldom epistemically isomorphic to axiomatic-deductive models/systems, and even if they were, we still have to argue for the external validity of the conclusions reached from within these epistemically convenient models/systems. Causal evidence generated by statistical/econometric procedures may be valid in 'closed' models, but what we usually are interested in, is causal evidence in the real target system we happen to live in.

Guy Orcutt once – according to (Leamer 1983, 31) -- said that "doing econometrics is like trying to learn the laws of electricity by playing the radio" Advocates of econometrics want to have deductively automated answers to fundamental causal questions -- but to apply 'thin' methods we have to have 'thick' background knowledge of what's going on in the real world, and not in idealized models.

Econometric forecasting

As Oskar Morgenstern noted in his classic *Wirtschaftsprognose: Eine Untersuchung ihrer Voraussetzungen und Möglichkeiten*, (Morgenstern, 1928) economic predictions and forecasts amount to little more than intelligent guessing. Making forecasts and predictions obviously isn't a trivial or costless activity, so why then go on with it?

The problems that economists encounter when trying to predict the future really underline how important it is for social sciences to incorporate Keynes's far-reaching and incisive analysis of induction and evidential weight in his seminal *Treatise on Probability* (Keynes, 1921).

According to Keynes, we live in a world permeated by unmeasurable uncertainty – not quantifiable stochastic risk – which often forces us to make decisions based on anything but 'rational expectations.' Keynes rather thinks that we base our expectations on the confidence or 'weight' we put on different events and alternatives. To Keynes, expectations are a question of weighing probabilities by 'degrees of belief,' beliefs that often have preciously little to do with the kind of stochastic probabilistic calculations made by the rational agents as modelled by 'modern' social sciences. And often we "simply do not know.'

How strange that social scientists and mainstream economists, as a rule, do not even touch upon these aspects of scientific methodology that seem to be so fundamental and important for anyone trying to understand how we learn and orient ourselves in an uncertain world. An educated guess on why this is a fact would be that Keynes's concepts are not possible to squeeze into a single calculable numerical 'probability.' In the quest for measurable quantities, one puts a blind eye to qualities and looks the other way.

So why do companies, governments, and central banks, continue with this more or less expensive, but obviously worthless, activity?

A part of the answer concerns ideology and apologetics. Forecasting is a non-negligible part of the labour market for (mainstream) economists, and so, of course, those in the business do not want to admit that they are occupied with worthless things (not to mention how hard it would be to sell the product with that kind of frank truthfulness). Governments, the finance sector and (central) banks also want to give the impression to customers and voters that they, so to say, have the situation under control (telling people that next year x will be 3.048 % makes wonders in that respect). Why else would anyone want to pay them or vote for them? These are sure not glamorous aspects of economics as a science, but as a scientist, it would be unforgivably dishonest to pretend that economics doesn't also perform an ideological function in society.

Econometric testing

Debating econometrics and its shortcomings one often gets the response from econometricians that "ok, maybe econometrics isn't perfect, but you have to admit that it is a great technique for empirical testing of economic hypotheses." But is econometrics — really — such a great testing instrument?

Econometrics is supposed to be able to test economic theories. But to serve as a testing device you have to make many assumptions, many of which themselves cannot be tested or verified. To make things worse, there are also only rarely strong and reliable ways of telling us which set of assumptions is to be preferred. Trying to test and infer causality from (non-experimental) data you have to rely on assumptions such as disturbance terms being 'independent and identically distributed'; functions being additive, linear, and with constant coefficients; parameters being' 'invariant under intervention; variables being 'exogenous', 'identifiable', 'structural and so on. Unfortunately, we are seldom or never informed of where that kind of 'knowledge' comes from, beyond referring to the economic theory that one is supposed to test. Performing technical tests is of course needed, but perhaps even more important is to know — as David Colander (2019, 340) puts it — "how to deal with situations where the assumptions of the tests do not fit the data."

That leaves us in the awkward position of having to admit that if the assumptions made do not hold, the inferences, conclusions, and testing outcomes econometricians come up with simply do not follow from the data and statistics they use.

The central question is 'how do we learn from empirical data?' Testing statistical/econometric models is one way, but we have to remember that the value of testing hinges on our ability to validate the — often unarticulated technical — basic assumptions on which the testing models build. If the model is wrong, the test apparatus simply gives us fictional values. There is always a strong risk that one puts a blind eye to some of those non-fulfilled technical assumptions that actually make the testing results — and the inferences we build on them — unwarranted.

Haavelmo's probabilistic revolution gave econometricians their basic framework for testing economic hypotheses. It still builds on the assumption that the hypotheses can be treated as hypotheses about (joint) probability distributions and that economic variables can be treated as if pulled out of an urn as a random sample. But as far as I can see economic variables are nothing of that kind.

It is still difficult to find any hard evidence that econometric testing uniquely has been able to 'exclude a theory'. As Renzo Orsi (1993, 365) put it: "If one judges the success of the discipline on the basis of its capability of eliminating invalid theories, econometrics has not been very successful."

The econometric illusion

The processes that generate socio-economic data in the real world cannot just be assumed to always be adequately captured by a probability measure. And, so, it cannot be maintained that it even should be mandatory to treat observations and data — whether cross-section, time series or panel data — as events generated by some probability model. The important activities of most economic agents do not usually include throwing dice or spinning roulette wheels. Data-generating processes — at least outside of nomological machines like dice and roulette wheels — are not self-evidently best modelled with probability measures.

When economists and econometricians — often uncritically and without arguments — simply assume that one can apply probability distributions from statistical theory to their own area of research, they are skating on thin ice. If you cannot show that data satisfies all the conditions of the probabilistic 'nomological machine,' then the statistical inferences made in mainstream economics lack sound foundations. The rigour and precision sought with mathematical statistical instruments has a devastatingly important trade-off: the higher the level of rigour and precision, the smaller the range of real-world application.

Statistical — and econometric — patterns should never be seen as anything other than possible clues to follow. Behind observable data, there are real structures and mechanisms operating, things that are — if we really want to understand, explain and (possibly) predict things in the real world — more important to get hold of than to simply correlate and regress observable variables.

Using formal mathematical modelling, mainstream economists sure can guarantee that the conclusions hold given the assumptions. However, the validity we get in abstract model worlds does not warrant transfer to real-world economies. Validity may be good, but it is not enough.

Mainstream economists are proud of having an ever-growing smorgasbord of models to cherry-pick from (as long as, of course, the models do not question the standard modelling strategy) when performing their analyses. The 'rigorous' and 'precise' deductions made in these closed models, however, are not in any way matched by a similar stringency or precision when it comes to what ought to be the most important stage of any economic research — making statements and explaining things in real economies. Although almost every mainstream economist holds the view that thought-experimental modelling has to be followed by confronting the models with reality — which is what they indirectly want to predict/explain/understand using their models — they then all of a sudden become exceedingly vague and imprecise. It is as if all the intellectual force has been invested in the modelling stage and nothing is left for what really matters — what exactly do these models teach us about real economies.

No matter how precise and rigorous the analysis, and no matter how hard one tries to cast the argument in modern mathematical form, they do not push economic science forwards one single iota if they do not stand the acid test of relevance to the target. Proving things 'rigorously' in mathematical models is not a good recipe for doing an interesting and relevant economic analysis. Forgetting to supply export warrants to the real world makes the analysis an empty exercise in formalism without real scientific value. In the realm of true science, it is of little or no value to simply make claims about a model and lose sight of reality.

To have valid evidence is not enough. What economics needs is sound evidence. The premises of a valid argument do not have to be true, but a sound argument, on the other hand, is not only valid but builds on premises that are true. Aiming only for validity, without soundness, is setting the economics aspiration level too low for developing a realist and relevant science.

Conclusion

Contemporary economics -- still -- focuses on studying what happens in abstract and unrealistic models. A deeper study of underlying causal mechanisms in the economy could make economic science more realistic. However, when faced with the monumental gap between empirical data and models, mainstream economists often resort to one of their four favorite strategies to immunize the models against facts:

1. Treat the model as an axiomatic system, which necessarily gives rise to pure logical tautologies.
2. Use unspecified *ceteris paribus* assumptions, which give each model assertion an unlimited 'alibi'.
3. Limit the applicability of the model to spatio-temporal problems where the assumptions/axioms are valid.
4. Leave the application of the model open-ended, making empirical falsification of the model impossible.

From a scientific standpoint, the value of these types of rescue actions is equal to zero. The real challenge for an economic science worth its name should be to confront reality as it is, instead of conjuring away all kinds of 'problems' through the formulation of models and theories that are nothing more than sterile deductive-axiomatic systems.

Science should not be reduced to substanceless fictional storytelling. This has been going on for far too long in economics.

References

Albert, H. (2012) "Model Platonism. Neoclassical Economic Thought in Critical Light." *Journal of Institutional Economics* 8 (3) pp. 295-323.

Bhaskar, R. (1997) *A realist theory of science*. New ed. London: Verso.

Colander, D. (2019) "Introduction to symposium on teaching undergraduate econometrics." *The Journal of Economic Education*, 50 (4) pp. 337-342.

Duflo, E. (2017) "The Economist as Plumber." *American Economic Review* 107 (5) pp. 1-26.

Friedman, M. (1953) "The Methodology of Positive Economics." In *Essays in Positive Economics* Chicago: University of Chicago Press.

Goldthorpe, J. H. (2001) "Causation, Statistics, and Sociology." *European Sociological Review*, *17* (1), pp. 1–20.

Haavelmo, T. (1944) "The Probability Approach in Econometrics." *Econometrica* 12 pp. iii–115.

Keynes, J. M. (1921) *Treatise on Probability* London: Macmillan.

Keynes, J. M. (1939) "The League of Nations Professor Tinbergen's Method." *The Economic Journal* 49 (195) pp. 558–577.

Lawson, T. (1997) *Economics and reality* London: Routledge.

Leamer, E. (1983) "Let's Take the Con Out of Econometrics." *American Economic Review 73* (1), pp. 31–43.

Morgenstern, O. (1928) *Wirtschaftsprognose: Eine untersuchung ihrer Voraussetzungen und Möglichkeiten* Vienna: Springer.

Nasir, M. & Morgan, J. (2023) "The methodological problem of unit roots: stationarity and its consequences in the context of the Tinbergen debate." *Ann Oper Res*.

Orsi, R. (1993) "Testing in econometrics: are economic theories testable?" *Journal of the Italian Statistical Society* 2, pp. 365–380.

Syll, L. P. (2016) "Microfoundations – on the use and misuse of theories and models in economics." In Madsen, M. Ove & Olesen, F. (eds.) *Macroeconomics after the financial crisis: a post-Keynesian perspective* London: Routledge.

Syll, L. P. (2023) "Deduction, Induction and Abduction." In *Routledge Handbook of Macroeconomic Methodology*, London: Routledge.

Author contact: lars.palsson-syll@mau.se

SUGGESTED CITATION:
Lars P. Syll, "Mainstream economics – the poverty of fictional storytelling", *real-world economics review*, issue no. 103, 31 March 2023, pp. 61-83, http://www.paecon.net/PAEReview/issue103/Syll

You may post and read comments on this paper at http://rwer.wordpress.com/comments-on-rwer-issue-no-103/

real-world economics review, issue no. 103
subscribe for free

Why do economists persist in using false theories?

Asad Zaman [International Islamic University, Pakistan]

It is easy to establish that many of the core theories used by Economists are false. For example, there is overwhelming empirical evidence against the theory of utility maximization; for a survey of this evidence, see Zaman and Karacuka (2011). Similarly, Romer (2016) documents how leading monetary economists persist in believing that monetary policy does not affect the real economy, despite very strong empirical evidence to the contrary. This failure of economic theory became obvious to all when economists failed to foresee the Global Financial Crisis (GFC) of 2007. Worse, leading economists confidently predicted continued prosperity, and dismissed warnings of trouble in financial markets After the crisis, many leading economists and practitioners realized that there were fundamental flaws in the structure of mainstream economic theories – for a collection of quotes, see "Quotes Critical of Economics".

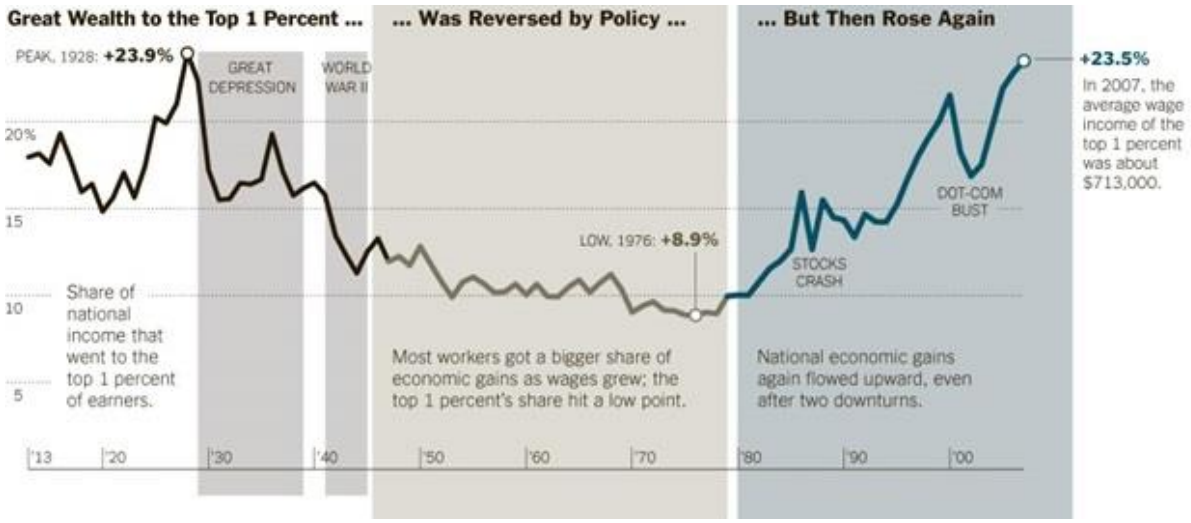
The question we wish to address is deeper than a one-time failure of economic theory. The same models which failed spectacularly in the GFC continue to be used after the crisis. The same models of consumer behavior overwhelmingly refuted by behavioral economists continue to be exposited in microeconomics textbooks, and taught to unsuspecting undergraduates around the globe. Central Banks around the globe continue to make monetary policy decisions on the basis of models known to be false. In a paper with the revealing title "Monetary Policy Without A Working Theory of Inflation", Daniel Tarullo, former head of the Federal Reserve, writes that " We do not, at present, have a theory of inflation dynamics that works sufficiently well to be of use for the business of real-time monetary policy-making". Around the world, Central Banks continue to raise interest rates to fight inflation, while the data overwhelmingly contradicts this causal link – see "Do High Interest Rates Reduce Inflation? A Test of Monetary Faith". So we can repeat the question of the title: Why do economists continue to use theories, even though they are well aware that empirical evidence is in strong conflict with these theories?

To answer this deeper question, we must dig deeper into the nature of economic theory itself. What is the function of economic theory, if it is not to learn the truth about how the economic system works? Once we explore economic theories within the historical contexts in which they arose, the answer becomes blindingly clear: economic theories serve to protect the interests of those in power. We provide a three examples of this below; for more, see ET1%: Blindfolds Created by Economics.

The Marginal Product of Capital: All economics textbooks argue that the returns to labor and capital (wages and interest) are determined by the technology, encapsulated in the production function, and the operation of competitive markets. None of them mention that this is a concealed moral argument to counter Marx's claim that Capitalists exploit labor. Since the returns to capital and labor are determined by the workings of the market mechanism in a symmetric way, with exactly the same mathematical form, we can conclude that both parties receive just compensation for their input to production. Both capital and labor are paid in proportion to their contribution to the productive process. Once we realize that a moral argument is being made, it becomes possible to counter this on several different grounds. Chapter 8 of Hill and Myatt's *Anti-Textbook of Microeconomics* provides a thorough discussion. One

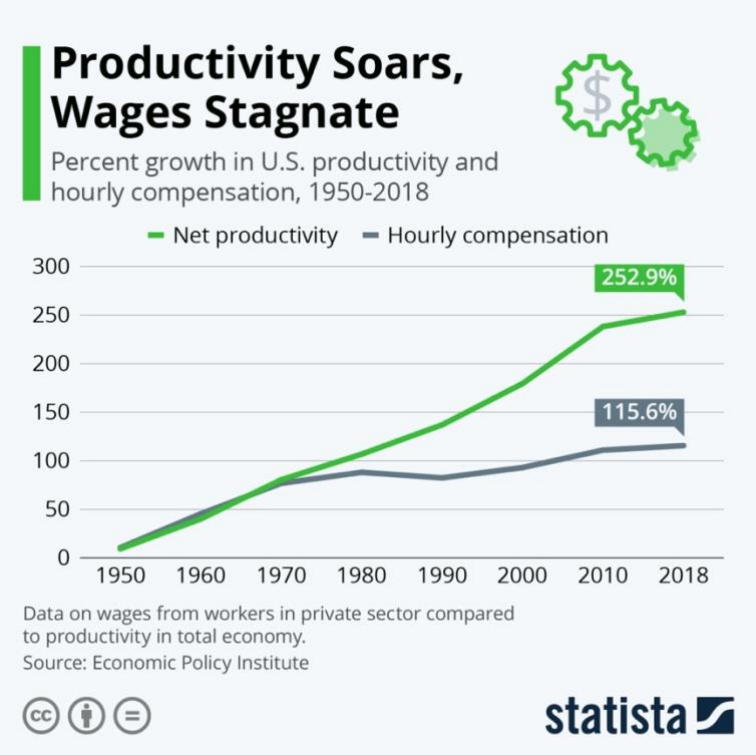
line of argument comes from the strong empirical evidence that most firms set prices, which shows that markets are not competitive. If the capitalists are able to set prices for their goods, then they can also set wages to be exploitative. A second line of argument comes from the mathematics. Production functions where the two marginal products (to capital and labor) exceed the total product, are easy to find. In such cases, it is technically impossible to pay the marginal product to both factors. Why do textbooks never mention such cases, and confine discussion to a simple special case where the two marginal products exactly equal the total product? A third line of argument comes directly from moral philosophy. Laborers labor to earn their wage. The capitalist "owns" capital to earn his reward – does ownership justify payment in the same way as working? One must look beyond all of these details to the bigger picture: A moral argument justifying payments to capitalists, and countering Marx's charge of exploitation, is being made in mathematical disguise.

The Keynesian Revolution and the Monetarist Counter-Revolution: The intimate connection between economic theory and political power is clearly illustrated by the rise and fall of Keynesian Economics in the 20th Century. Confidence generated by theories glorifying the workings of a market economy led leading economists to predict permanent prosperity, just prior to the Great Depression of 1929. After the crash, Keynes set out to resolve the most glaring contradiction between economic theory and reality. While economic theory maintains that free markets automatically eliminates unemployment, the Great Depression created high unemployment which persisted for more than a decade. Keynesian theory recognized this failing of free markets, and placed responsibility for creating full employment on the government. Application of Keynesian theory led to a period of unprecedented prosperity in Europe and USA following the 2nd World War. However, there was a snake in the Garden of Eden: the wealth share of the top 1% declined precipitously between 1930 and 1980:



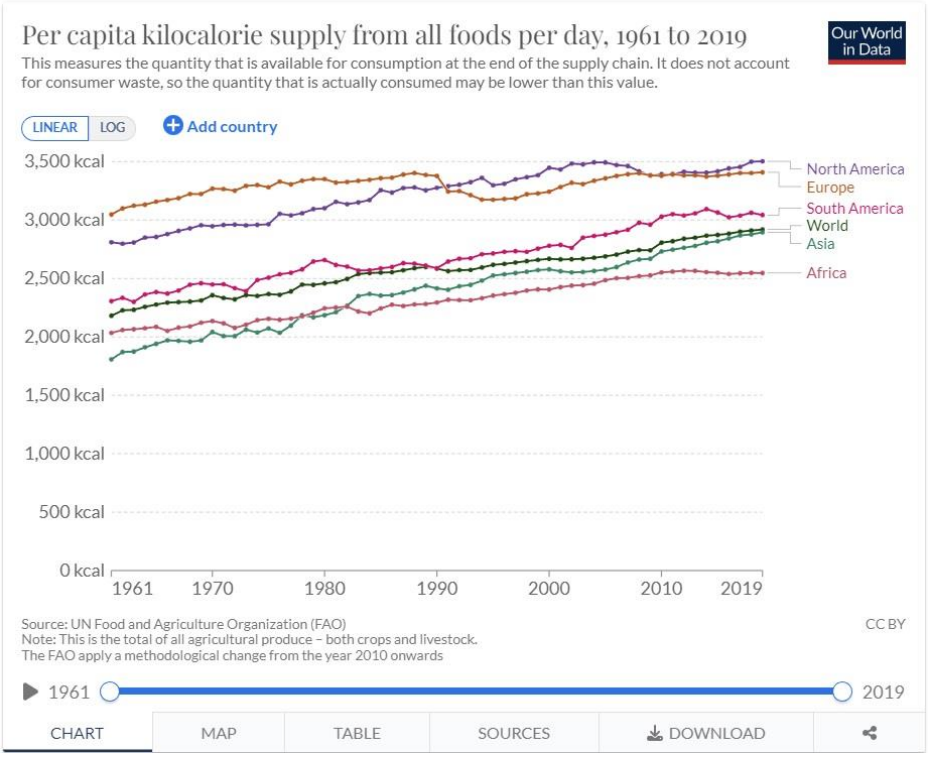
The top 1% fought back by a well-thought out multi-dimensional plan to reverse this decline in their wealth shares; details of this planning are available from Naomi Klein's Shock Doctrine, and Alkire and Ritchie Winning Ideas: Lessons from Free Market Economics. A central element of this plan, implemented in the Reagan-Thatcher era, was the rejection of Keynesian economics and a return to the same pre-Keynesian ideas that had been proven wrong by the Great Depression of 1929. Modern textbooks of labor theory continue to teach that free markets eliminate unemployment, blithely ignoring the massive amounts of empirical evidence against this proposition. Chicago school economists argued that government interventions to create full employment bring about short term increases in employment, which are reversed in the long run. Furthermore, such interventions inflict great costs upon

the economy in the form of high inflation. Central Banks responded by dropping the goal of reducing unemployment, and shifting policy focus to fighting inflation only. The result was a long period of economic stagnation, with high unemployment, which weakened power of labor force and enabled capitalist exploitation, reflected in the rapid rise of the wealth share of the top 1%. Another graph which shows that the productivity increased a lot, but the wealthy captured the lion's share of these gains, while the labor share remained nearly constant, is given below:



This clearly demonstrates why economics textbooks stick to the theory that free markets create full employment, when they obviously do not (see: 70 years of failure by economists to understand the labor market). Allowing unemployment to exist, and preventing the government from intervening to eliminate it, permits capitalists to exploit labor to the hilt, appropriating all gains from increasing productivity, and denying labor any share of the increasing profits.

The Scarecrow of Scarcity: As detailed in my paper on The Normative Foundations of Scarcity, the foundations of economics were shifted from an approach focusing on material welfare, to an approach based on scarcity. When we ask "Why?" we find the same answer: the concept of scarcity is designed to conceal the wealth of the rich and protect it against the claims of the poor. This is a continuation of a strategy adopted by Malthus, who argued, purely from his imagination and without any data, that poverty was due to the high fertility rate of the poor, and giving them food would only increase this rate of growth and be counterproductive. The proponents of scarcity argue that the reason that there are over a billion people living below the poverty line on the planet is because of the scarcity of resources to feed them. However the data on food per capita contradicts this view:



The data shows that in all continents, including Africa, the per capita kilocalories supplied by food has been increasing, and is above minimal requirements. So poverty does not exist because there is not enough food for all, it is because planetary resources are concentrated in the hands of a few rich people. As Gandhi observed: "There is enough for everyone's need, but not enough for everyone's greed". The obvious solution to the problem of poverty lies in the redistribution of wealth. But, since this would harm the interests of the wealthy, this line of thinking is actively discouraged. "Nobel Laureate" Lucas states that:

> "Of the tendencies that are *harmful* to sound economics, the *most* seductive, and in my opinion the *most* poisonous, is to focus on questions of *distribution*. The potential for improving the lives of poor people by finding different ways of distributing current production is nothing compared to the apparently limitless potential of increasing production."

While many different lines of research have converged on the truth that elimination of poverty requires redistribution of wealth concentrated in the hands of a tiny minority, economists stubbornly refuse to acknowledge this, and insist that we must create more growth and acquire more wealth in order to be able to feed the poor.

Concluding Remarks: Richard Feynman argued that the defining characteristic of scientific theories is that they are rejected and revised when they conflict with experimental evidence: "It doesn't matter how beautiful your **theory** is, it doesn't matter how smart you are. If it doesn't agree with **experiment**, it's **wrong**". These three examples above, as well as a host of others not discussed, show that the defining characteristic of textbook economic theory is that it remains the same regardless of how much empirical evidence accumulates against it. A closer examination of progress in economic theory shows clearly that the core has remained the same, while the evolution and progress has been in the complex defense mechanism devised to protect these core theories against mountains of accumulating conflicts with empirical realities. Regarding these tendencies, Romer writes that the economists' "*dismissal of*

fact goes so far beyond post-modern irony that it deserves its own label. I suggest "post-real." While it is impossible to understand the evolution of economic theory as a progressive increase in understanding the complex economic reality, it is easy to understand all resistance to change when it is viewed as an ideology designed to protect the interests of the wealthy and powerful.

Author contact: asad.zaman@alumni.stanford.edu

SUGGESTED CITATION:
Asad Zaman, "Why do economists persist in using false theories?", *real-world economics review*, issue no. 103, 31 March 2023, pp. 84-88, http://www.paecon.net/PAEReview/issue103/Zaman

You may post and read comments on this paper at http://rwer.wordpress.com/comments-on-rwer-issue-no-103/

real-world economics review, issue no. 103
subscribe for free

Revisiting the Principles of Economics through Disney

Junaid Jahangir [MacEwan University, Canada]

Abstract
The objective in this paper is to revisit the principles of economics taught at the principles level. Many ECON 101 students end up with the dogma of market fundamentalism or economism. Large scale inertia prevents a complete overhaul of ECON 101. There are also concerns with both mainstream and heterodox economists dismissing alternative perspectives and with the chalk and talk method of instruction. Thus, the focus in this paper is on comparatively viewing neoclassical and heterodox perspectives and illustrating economic principles through Disney video clips. Such a strategy elicits student interest due to familiarity and connection with childhood memories and helps with recall in the age of information overload. Moreover, the instructor preparation time and technological requirements of such an approach are minimal. Overall, with pluralist perspectives and Disney clips, the twin objectives of upholding nuance and retaining student interest are achieved.

Keywords: ECON 101; Disney; heterodox economics; ten principles
JEL Codes: A22
Disclosure Statement: There are no competing interests to declare.

1. Introduction

At the 56th annual conference of the Canadian Economics Association, Hill (2022) focused on the choice between "nonstandard" textbooks and "supplementary critical commentaries" to standard neoclassical textbooks in teaching microeconomics principles. In his own co-authored book, Hill and Myatt (2021) present the standard textbook perspective on economics principles followed by the "anti-textbook" perspective as a critical counterpart. In doing so, they seemingly adopt the pluralist perspective to introduce students to both neoclassical and heterodox perspectives. This approach is parallel to that of Lee (2010) who argued that undergraduate students should be exposed to different economic theories to allow them to make up their own minds, that it is important to have a deep understanding of various theories, and that neoclassical economics must be studied to learn about the historical development of standard textbook theory. While some instructors would like to use a heterodox textbook and completely overhaul the way ECON 101 is taught, Colander (2015) reminds us of large-scale inertia, as many instructors have invested significant human capital in the neoclassical paradigm.

My own orientation is to retain a dialogue with the neoclassical paradigm, instead of a radical overhaul in teaching ECON 101. This is partly due to institutional requirements to maintain a certain level of homogeneity across ECON 101 sections taught by different instructors, and partly due to my understanding that students should be facilitated to make up their own minds on addressing economic issues. I am also new towards discovering heterodox economics. However, I have begun introducing

critiques of the neoclassical paradigm on the principles of economics in both ECON 101 and in a newly designed ECON 357 topics course called Humanistic Economics. This course is based on Komlos (2019) and critiques the standard treatment of various topics including equilibrium, minimum wage, tariffs, free trade, and inequality. Thus, instead of the binary set up by Hill (2022), I have introduced critical commentary of standard textbook theory in ECON 101 and exposed students to heterodox economics in a newly designed course. However, of all the topics, I have spent more time on critically evaluating the ten principles of economics, as delineated by Mankiw, Kneebone, and McKenzie (2020), which is used as a standard textbook in our economics department.

The focus on the principles of economics is warranted, as many ECON 101 students neither major in economics nor take higher level classes in the subject. They are therefore left with the dogma of market fundamentalism or economism (Komlos, 2021a; Hill and Myatt, 2021, pp. 2-3). These supposedly "axiomatic" principles include the notions of trade-offs, opportunity costs, rationality of individuals, incentives, free trade, efficient allocation of markets, limited government intervention, productivity as the chief determinant of living standards, inflation as a monetary phenomenon, and the short run trade-off between inflation and unemployment. The emphasis on these principles is also warranted given the critiques of ECON 101 led by both student groups (Earle, Moran, and Ward-Perkins, 2017) and faculty (Skidelsky, 2020; Bowles and Carlin, 2020; Komlos, 2019; Reardon et al., 2018) especially in the aftermath of the 2008 financial crisis, which the neoclassical paradigm failed to explain. Amongst these critiques, Goodwin (2014), Campbell et al. (2019) especially focus on the ten principles whereas Hill and Myatt (2021) focus on the subset of microeconomics related principles.

However, Colander (2016) has argued that the principles can be taught with nuance. This alludes to the idea of not throwing the baby out with the bath water. Similarly, Hill and Myatt (2021) have argued that the neoclassical paradigm is "remarkably malleable" so that it is "capable of transforming" and "shedding" its "unappealing" features (p. 1). What this means is that when students are introduced to the principles, it should be done in a nuanced manner that captures the best of both the neoclassical and heterodox paradigms. This would limit the impact of market fundamentalism or economism on ECON 101 students that do not pursue higher economics courses. Moreover, while it is important to facilitate a nuanced understanding, it is equally important to retain student interest and make economic content relatable. This is where the Becker and Watts (1996) critique of the chalk and talk method becomes relevant. Indeed, there has been a surge of literature that involves the use of tools including animated cartoons to teach economics (Al-Bahrani et al., 2016). At the ECON 101 level, Knudsen and Duncan (2018) showcase Disney animations to teach economics principles and concepts. They argue that popular culture catches student interest in a way that traditional pedagogies cannot. More recently, Mandzik (2022) has illustrated economics principles and concepts through fairy tales including Disney's The Little Mermaid, Cinderella, and Aladdin. Such methods are used to make economic content more relatable, increase student interest and their retention of economic ideas (Acchiardo et al., 2017).

However, this literature predominately focuses on neoclassical economics with some exceptions including Jahangir (2021; 2022). Although, none of them focus on all ten principles comprehensively. Thus, the focus in this paper is to revisit the ten "axiomatic" principles by pairing them with Disney video clips to illustrate both the neoclassical and heterodox perspectives. In doing so the twin objectives of upholding nuance and retaining student interest are achieved. Moreover, the instructor preparation time, class time, and technological requirement of showing such clips are minimal but the benefit is high in terms of eliciting student interest, engendering class discussion, and creating memorable lessons. Additionally, no prior familiarity with entire feature length Disney animations and movies is necessary to lead class discussions. Thus, this paper is structured in five sections including the Introduction. In Section 2, the ten principles are addressed especially considering the critiques by Goodwin (2014), Campbell et al. (2019), Reardon et al. (2018), Komlos (2019), and Wray (2022) and each principle is

paired with a Disney video clip to elicit student interest and generate class discussion. In Section 3, alternative principles are highlighted considering Campbell et al. (2020), Komlos (2021a), and Hill and Myatt (2021) amongst others. In Section 4, student feedback from my Humanistic Economics class is presented where I spent a considerable time addressing these principles through both the neoclassical and heterodox paradigms. This is followed by concluding remarks in Section 5.

2. Revisiting the Principles of Economics

The ten supposedly "axiomatic" principles, as delineated by Mankiw, Kneebone, and McKenzie (2020), include the notions of trade-offs, opportunity costs, rationality of individuals, incentives, free trade, efficient allocation of markets, limited government intervention, productivity as the chief determinant of living standards, inflation as a monetary phenomenon, and the short run trade-off between inflation and unemployment. These principles are revisited systematically as follows.

2.1 Trade-Offs (There is always a choice)

In neoclassical economics, the first principle is often illustrated by the choices between consumption and saving, growth and environment, and equity and efficiency. However, this principle can be used to showcase that rising military expenditures or instituting corporate tax cuts come at the expense of reduction to healthcare and education. Thus, society faces a trade-off or choice between corporate handouts and social welfare. Such an understanding of neoclassical economics aligns one closer to heterodox perspectives. However, in a world that increasingly questions binary options, students may allude to the Green New Deal to push the point that we can have more employment and a cleaner environment as we restructure the economy towards the renewable energy sector. As such, standard ECON 101 thinking faces challenge from out of the box thinking with the non-binary. Additionally, Campbell et al. (2019) argue that the principle of trade-offs does not account for situations where there is no true choice. Such is the case between taking a dangerous job or facing eviction. In other words, there isn't much of a choice between the devil and the deep blue sea. Thus, recognizing non-binary thinking that rejects binary options and equally the idea that true choice does not always exist, both allow to effectively nuance the principle of trade-offs. In terms of the Disney animation relevant to this principle, a short video clip from Beauty and the Beast can be shared to highlight the point that Belle does not have much of a choice in freeing her father Maurice when she decides to take his spot as the Beast's prisoner.

Figure 1: Belle takes her father's spot

Credit: DISNEY
Video Clip: https://www.youtube.com/watch?v=1tVQ-o8PleU
Duration: 4:02 minutes

Discussion: The strategy would be to show this clip as a hook to draw students into a discussion on the absence of choice in real life scenarios in conjunction with the issues of exploitation and power. The analogy can be made that just as the Beast is in a position of power that renders Belle no reasonable choice, so too is the case for the working-class poor who worked as frontline grocery and sanitation workers during the pandemic and who were at a greater risk of contracting COVID through public transport and proximity to other workers in factories. Such workers did not have the choice to take their work home or to risk leaving their jobs. Thus, a nuanced understanding on the absence of choice in real life situations can be provided apart from the idea that societies face a trade-off between corporate handouts and social welfare.

2.2 Opportunity Costs (There is no such thing as a free lunch)

The second principle is closely tied to the first, as it emphasizes the value of the next best alternative foregone. Known as the opportunity cost of the option undertaken, this principle is used to highlight that there is no such thing as a free lunch. Like the first principle, it can be used to reiterate that the opportunity cost of corporate tax cuts is the foregone healthcare and education spending. Viewed as such, this principle can be marshalled for the goals prized within heterodox circles. However, Campbell et al. (2019) critique that free lunches exist in the context of "forcible appropriation of resources". This alludes to the notion that advanced economies were able to build their wealth on the back of slave labour and resources expropriated from developing countries through colonialism. In a similar vein, Wray (2022) argues in the context of Modern Monetary Theory that a free lunch exists when unemployed resources can be hired to build capacity to generate current and future income and employment, and that it is a policy failure not to do so (pp. 111, 112). In terms of the Disney animation relevant to this principle, a short video clip from Pocahontas can be shared to underscore the point on free lunch when Governor Ratcliffe as a settler from England claims the land and riches of the natives in the name of King James thereby naming the land Jamestown.

Figure 2: Governor Ratcliffe claims Jamestown

Credit: DISNEY
Video Clip: https://www.youtube.com/watch?v=2yJzkO9fVh8
Duration: 1:42 minutes

Discussion: The strategy would be to show this short clip as a hook to draw students into a discussion on the existence of a free lunch through forcible appropriation of slave labour and resources in the context of colonialism. The connection can be made between Governor Ratcliffe claiming land and resources in the name of King James and the modern-day equivalent of bonded labour in developing countries where generations are born into debt that never seems to be paid off. Additionally, the issue of debt trap can be discussed where ports and large infrastructure have been taken over by China when developing African and Asian countries fail to repay their debt, as in the case of the Hambantota Port in Sri Lanka. Thus, as in the case of the first principle, the issues of power and exploitation can be raised that provide a nuanced understanding on free lunch beyond the simple lessons of facing choices and opportunity costs.

2.3 Rational people think at the margin (Marginal analysis)

The third principle is based on assuming rationality, which is defined as having the ability and information to make decisions (preferences are complete), making decisions consistently (preferences are transitive), and that more is preferred to less. A rational individual therefore undertakes a cost-benefit analysis at the margin in their decision-making process. The emphasis on the margin allows the individual to ignore sunk costs in decisions such as deciding to exit a movie theater or leaving food at a fancy restaurant after having paid for them respectively. This principle can be invoked to show that we can get out of sub optimal situations by treating past investments as sunk costs. In other words, we have the power to not be defined by past decisions or misfortunes and create a new path. Similarly, Hill and Myatt (2021) indicate that the assumption of rationality does not mean that individuals are only motivated by self-interest but that they "have consistent preferences" so that rationality would incorporate altruism in that the utility of one individual may depend on the welfare of others (p. 14). While such framing would draw this principle closer to heterodox perspectives, this principle has been subjected to multiple critiques.

Hill and Myatt (2021) critique the assumption of rationality in explaining individuals getting addicted to hard drugs (p. 25), a point conceded by Mankiw, Kneebone, and McKenzie (2020) when they state that

drug addicts are not looking after their best interests (pp. 150-151). Similarly, Campbell et al. (2019) argue against the assumption of rationality or reducing decision making to a mechanical "first derivative condition" by alluding to imperfect information and the limited cognitive abilities of individuals. Goodwin (2014) argues against the rationality of individuals by mentioning that we are swayed by advertisements to buy products that make us worse off like fried meat and soda drinks, and that we are manipulated through framing, as we are more likely to buy a good at $3.99 than at $4.00 despite the nearly identical price. According to Komlos (2019), given differential education and cognitive abilities and given the cost of acquiring information, sellers engage in opportunistic behaviour and exploit individuals through fine print and information overload (p. 59, 93). He argues that neither consumers nor firms engage in marginal analysis (p. 110). Moreover, instead of maximizing utility, consumers make decisions based on rules of thumb, that is, satisficing instead of optimizing, and that firms use the mark up rule to obtain the selling price and use heuristics instead of marginal product to determine wages (pp. 60-63, 100, 116). Finally, George (2022) critiques marginal analysis and the mathematics of profit maximization by arguing that a firm cannot solely focus on marginal costs and ignore fixed costs when setting its price.

Overall, the multiple critiques on rationality and marginal analysis include the limits of rationality in explaining drug addiction, the manipulation of corporations through advertisements, framing of prices, exploitation through fine print, the use of heuristics in decision making, satisficing instead of utility maximizing, and incorporating fixed costs in decision making instead of setting prices based on just marginal costs. In terms of the Disney animation relevant to this principle, a short video clip from Snow White and the Seven Dwarfs can be shown to highlight the point that the evil queen disguised as an old hag is able to manipulate a naive Snow White to take a bite of the poisonous apple because the cognitive abilities of the latter do not allow her to see beyond her desire of having a "happily ever after" with Prince Florian.

Figure 3: Snow White gets manipulated

Credit: DISNEY
Video Clip: https://www.youtube.com/watch?v=Kl3qXO7D9vU
Duration: 2:58 minutes

Discussion: The strategy would be to show this short clip as a hook to draw students into a discussion on the tactics of corporations to get consumers to buy products they don't need through advertisements that project care or appeal to emotions and by framing prices as discount offers. The connection can be made between how the evil queen projects herself as an "old granny" who preys on a young Snow White by pretending to care about her wishes and the tactics of corporations as they tailor their

advertisements to appeal to emotions or that have good looking youth having fun only to sell unhealthy junk food including fries and soda. Thus, a refined understanding on this principle can be provided in that individuals do not always act rationally and are therefore susceptible to the tactics of corporations beyond the simple lesson that individuals are only motivated by self-interest.

2.4 People respond to incentives

The fourth principle that people respond to incentives is not limited to monetary incentives and its main thrust is to show that we get the opposite of what we wanted. For instance, Mankiw, Kneebone, and McKenzie (2020) show how in response to the seat belt laws, more accidents took place as people felt safer to drive faster (pp. 6-9). Another example of unintended consequences is of the Cobra Effect where a British officer in colonial India offered financial incentives to get rid of the snakes only to find that the local population started breeding snakes for the monetary reward (Dubner, 2012). Therefore, recognition of this principle allows policy makers to be mindful of the unintentional consequences that may result based on their otherwise well-intentioned policies. However, Goodwin (2014) critiques that viewing the principle in such a manner maintains the status quo for it suggests that any government intervention to improve road safety would be futile. He adds that financial incentives crowd out intrinsic motivation as in the case of blood donations where people gave less blood when it is about earning a few dollars than when it is viewed as a noble and selfless act. Hill and Myatt (2021) echo this critique when they state that financial incentives can "reduce altruistic behaviour" and allude to the case where parents started arriving late to pick up their nursery children after a small fine was instituted, as the payment reduced the guilt on arriving late (p. 27). Similarly, Reardon et al. (2018) critique the idea that people respond to incentives because they may not be rational, have established habits or brand loyalty, may be desperate or have little market power compared to the sellers (p. 53).

Overall, the critiques of this principle indicate that financial incentives can crowd out intrinsic motivations and altruistic behaviour, and that people may not respond to incentives because of factors including brand loyalty, entrenched habits, or other considerations. In terms of the Disney animation relevant to this principle, a short video clip from Mulan can be shown to highlight the point that despite the Emperor's offer of a seat in his council that would bring rank and prestige, Mulan is not swayed and decides instead to go back to her family.

Figure 4: Mulan rejects the Emperor's offer

Credit: DISNEY
Video Clip: https://www.youtube.com/watch?v=4i4TLmF73dk
Duration: 3:29 minutes

Discussion: The strategy would be to show this short clip as a hook to draw students into a discussion on the idea that people do not respond to financial or other incentives as is usually expected. The connection can be made between Mulan's decision to refuse the prestigious seat at the Emperor's council out of family considerations and the decision of many individuals to put family first before financial rewards and high profile job positions. A discussion can also be broached on how individuals are not swayed by financial incentives on energy efficiency investments because of entrenched habits and how a better way to motivate people would be through promoting values instead of throwing a few dollars. Thus, a comprehensive understanding can be provided so that students learn that people do not always respond to incentives or that financial incentives can crowd out intrinsic motivations just as they learn that offering incentives can have unintended consequences.

2.5 Trade can make everyone better off

The fifth principle that trade can make everyone better off is used to show that free trade allows access to cheaper, greater variety, and more goods. Moreover, restrictions on trade through tariffs, justified by arguments including the infant industry argument, lead to dead weight loss (inefficiency) and confer protection to politically powerful industries (Mankiw, Kneebone, and McKenzie, 2020, p. 209). However, Reardon et al. (2018) argue that free trade is not necessarily fair trade, that mass production is undertaken in countries with poor environmental standards and low wages, and that trade takes a heavy toll on the environment (pp. 296, 297). They mention that nations like U.K. and the U.S. started off as highly protectionist and only started preaching free trade after they become economic superpowers, as they needed unrestricted access to consumer demand in the global market (p. 224). On Ricardo's theory of comparative advantage, they highlight the unrealistic assumptions of his theory and argue that it was developed in a time without concern for environmental limits (pp. 310, 311). Finally, they state that many developing countries have not achieved sustainable economic development based on market liberalization and reduced government intervention, as pushed by global institutions like the IMF, WTO and the World Bank, and that free financial capital flows contribute to macroeconomic instability and unemployment (pp. 279, 288, 318). In a similar vein, Komlos (2019) argues for the welfare enhancing property of tariffs for even if consumers lose because of higher prices, these losses are

dispersed among many people, whereas the gains are concentrated among the smaller group of underemployed workers that get jobs (pp. 222, 225).

Overall, like the principle on rationality, this principle is contested on multiple grounds including the unrealistic assumptions of the theory used to justify free trade, the impact on the environment and workers' wages, the effect of free capital flows on macroeconomic instability and unemployment, and the welfare enhancing property of tariffs. In terms of the Disney animation relevant to this principle, a short video clip from The Little Mermaid can be shown to highlight the point that the free trade between Ariel and the sea witch Ursula makes Ariel worse off, as she gives up her voice in exchange for human legs and becomes susceptible to being forever enslaved under the terms of contract.

Figure 5: Ariel trades her voice

Credit: DISNEY
Video Clip: https://www.youtube.com/watch?v=Y0JoW27fxUw
Duration: 2:31 minutes

Discussion: The strategy would be to show this short clip as a hook to draw students into a discussion on the idea that free trade agreements can be detrimental for weaker parties like labour in advanced countries that face unemployment, exporters in developing countries that face unfavourable terms of trade, and governments facing balance of payment crises that are dissuaded by global institutions from pursuing full employment policies through fiscal policy. The connection can be made between Ariel's decision to give up her voice and sign an unfair contract, which makes her susceptible to being forever enslaved, and developing countries that sell their resources at unfavourable terms of trade, open themselves up for corporations that damage the environment, and become susceptible to balance of payment and currency crises with unfettered capital flows. Thus, through the analogy showcased with this clip, students can be offered a nuanced understanding on the detrimental impact of free trade apart from the touted benefits of greater variety, cheaper, and more goods.

2.6 Markets are usually a good way to organize economic activity

The sixth principle is essentially about free markets allocating resources efficiently in contrast to information requirements and planning delays in a centrally planned economy like the U.S.S.R. Hill and Myatt (2021) emphasize that markets do not require "expensive planning bureaucracy" or that people be "altruistically motivated" to maximize "society's net benefit" (pp. 14-15). However, even Mankiw, Kneebone, and McKenzie (2020) note that markets need government institutions to enforce property rights and that the "the invisible hand does not ensure that everyone has sufficient food, decent clothing, and adequate health care" (pp. 12-13). Similarly, Komlos (2019) argues that markets only work towards efficiency with appropriate government-provided laws and institutions and that bounded rationality, opportunistic behaviour, asymmetric information, monopolies and oligopolies, transaction costs, and pollution prevent markets from working efficiently (pp. 31, 49). Likewise, Reardon et al. (2018) reject the perfect competition model and question defining efficiency as allocating resources to produce the greatest output as opposed to sustainability or best utilizing workers' talents (p. 175). Finally, both Campbell et al. (2019) and Goodwin (2014) highlight that in cases like healthcare insurance, the government is better able to pool risks and negotiate with pharmaceutical companies owing to economies of scale, as opposed to the private sector.

Overall, the critiques of this principle indicate that markets do not maximize efficiency in the absence of government institutions, that the presence of various factors like bounded rationality, opportunistic behaviour, and monopolies and oligopolies impede market efficiency, that government provision is superior to the private sector in cases like healthcare insurance, and that efficiency should be replaced by sustainability as the overarching goal. In terms of the Disney animation relevant to this principle, a short video clip from Cinderella can be shown to highlight the point that it is royal proclamation (government decree) on inviting every eligible maiden to the ball that creates a level playing field for Cinderella to attend and without which, she would not even have hoped to be at the ball given the unfair treatment by market participants including her stepmother and stepsisters.

Figure 6: Cinderella's Invitation to the Ball

Credit: DISNEY
Video Clip: https://www.youtube.com/watch?v=IkFNEUrSD3w
Duration: 2:21 minutes

Discussion: The strategy would be to show this short clip as a hook to draw students into a discussion on the idea that government institutions and rules are necessary for efficient functioning of the market

in the first place. The connection can be made between the royal decree that allows every eligible maiden including Cinderella to attend the ball, despite the unfair treatment of her stepmother and stepsisters, and government rules that create a level playing field for individuals and firms to fairly compete in the market given powerful economic actors that resort to unfair tactics and exploitation. Thus, through the analogy showcased through this clip, students can be offered a nuanced understanding on the essential role of the government that facilitates effective functioning of the markets in the first place. The point is to understand that markets do not exist in vacuum and that they are dependent on institutions and rules provided by the government apart from acknowledging the standard lesson that centrally planned economies are marred by inefficient bureaucracies, delays, and informational constraints.

2.7 Governments can sometimes improve market outcomes

The seventh principle is on government intervention required in the case of market failure due to market power and externalities like pollution where markets do not exist for clean air. In general, neoclassical economics supports the use of patents to spur innovation and Pigouvian taxes to correct externalities. Although, it does not countenance government intervention for price ceilings (rent controls) or price floors (minimum wage) by alluding to deadweight loss (inefficiency) that arises because markets no longer remain in equilibrium. However, Reardon et al. (2018) critique that patents protect monopoly power and therefore do not spur but deter innovation, and that using Pigouvian taxes simply reinforces polluting behaviour and the mindset that a fee can always be paid (pp. 180, 204). They reject the notion of equilibrium as an ossified idea that rests on Newtonian physics and highlight the post-Keynesian defense of minimum wage, which increases the workers' standard of living and increases consumer demand that in turn supports employment (p. 135). Similarly, they justify rent controls because of "slum" landlords, who historically charged high rents for poor quality apartments (p. 214). Likewise, Komlos (2019) notes several examples of price gouging by pharmaceutical companies, which necessitates a cap on prices and argues that capping the salaries of most CEOs would not be inefficient, as their earnings are mostly rent (p. 119, 161). Additionally, Komlos (2021b) critiques that neoclassical theory ignores exploitation, treats consumer protection as superfluous, and rejects the idea that discriminating firms will be outcompeted in the market, as there may not exist sufficient non-discriminating firms that could withstand social pressures on racism. He adds that it was not markets but government regulation that ensured that Rosa Parks could sit in the front of the bus. Finally, Campbell et al. (2019) highlight that government intervention is essential for health and safety regulations.

Overall, this principle is critiqued on the grounds that government intervention is not just required to address market power and externalities but also safety regulation, anti-discrimination, and equity. Moreover, the neoclassical view of price ceilings and floors as causing inefficiency is countered by rejecting the ossified notion of equilibrium and alluding to the beneficial impact of minimum wage and price caps. In terms of the Disney animation relevant to this principle, a short video clip from Aladdin can be shown to highlight the point that it was the Sultan (the government) who had to overturn the discriminating law that would eventually allow a poor Aladdin to marry the Princess Jasmine, thereby facilitating upward social mobility.

Figure 7: The Sultan changes the law on marriage

Credit: DISNEY
Video Clip: https://www.youtube.com/watch?v=6aTUCi27Y0w
Duration: 3:45 minutes

Discussion: The strategy would be to show this short clip as a hook to draw students into a discussion on the idea that government intervention through regulation and laws is required for anti-discrimination and equity considerations. The connection can be made between the royal decree that allows a poor Aladdin to marry Princess Jasmine and government regulations on anti-discrimination and equity based on factors including race, gender, income status, and other considerations. Moreover, the discussion can be broadened to include health and safety regulations, workplace safety, minimum wage and price caps, all factors that warrant government intervention. Thus, through the analogy showcased with this clip, students can be offered a refined understanding that government role goes well beyond addressing market failure based on market power and externalities.

2.8 Productivity is the main determinant of living standards

The eighth principle is about attributing variation in living standards predominantly to differences in productivity, which is based on GDP per worker. Notwithstanding the limitations of the concept of GDP, Mankiw, Kneebone, and McKenzie (2020) state that "the growth rate of a nation's productivity determines the growth rate of its average income" so that improving living standards would warrant raising productivity (pp. 13, 14). However, Campbell et al. (2019) state that this principle is "not very insightful", as it does not consider the quality of life and the conditions of work. Campbell et al. (2020) go further to highlight that capitalists exploit workers by paying them less than the value they create and that the threat of unemployment limits workers' ability to obtain higher wages. Likewise, Goodwin (2014) critiques that high productivity is associated with the degradation of the environment. Thus, when productivity increases, living standards do not rise due to the negative impact of inequality and climate change. In a similar vein, Komlos (2021c) rejects equating living standards with productivity by arguing that "output does not translate automatically into well-being or happiness" and that "we no longer need an ever-increasing quantity of goods". Komlos (2019) mentions that from 1982 to 2016, productivity increased by 94% but compensation only increased by 40% in the U.S. (p. 112). Instead of blaming globalization or technological change, he argues that "firms took advantage of their power and paid workers far less than what they were worth" (p. 112). Thus, he rejects marginal analysis that equates wages to the value of the marginal product of labour.

Overall, the critiques of this principle emphasize that it ignores work conditions and quality of life, that rising productivity is associated with inequality and environmental degradation, that higher productivity does not translate to greater happiness, and that corporations exploit workers as productivity increased but real wages stagnated. In terms of the Disney animation relevant to this principle, a short video clip from Pinocchio can be shown to highlight how the puppeteer Stromboli makes money off Pinocchio's labour while giving him the odd coin, exploiting him to work in multiple shows, keeping him in a cage, and threatening to use him as firewood once he became too old to work.

Figure 8: Stromboli exploits Pinocchio's labour

Credit: DISNEY
Video Clip: https://www.youtube.com/watch?v=MKGCymtaiMk
Duration: 2:40 minutes

Discussion: The strategy would be to show this short clip as a hook to draw students into a discussion on the poor working conditions and low wages under capitalism. The connection can be made between Stromboli exploiting Pinocchio's labour with poor pay and work conditions and the living conditions of the working poor that work for large corporations in advanced economies and in places in the Middle East. The discussion can include places like Qatar and the U.A.E. where expatriate workers from developing countries face human rights violations and poor work conditions as they are packed in rooms with bunker beds and without any pension plan or old age security. Thus, through the analogy showcased through this clip, students can be offered a refined understanding that greater productivity does not always translate to higher living standards or happiness, as the principle ignores issues of work conditions, quality of life, and exploitation by powerful corporations.

2.9 Inflation is a monetary phenomenon

The ninth principle attributes inflation predominately to printing too much money. It is used to caution against Zimbabwe style hyperinflation under the Mugabe regime. The exposition on this principle is concise in Mankiw, Kneebone, and McKenzie (2020) compared to other principles. This lesson against printing money has been illustrated by Luccasen et al. (2011) who use clips from the animation Duck Tales to explain inflation and long run monetary policy apart from addressing other macroeconomic issues. However, Wray (2022) critiques this principle by alluding to supply side factors that cause inflation including oil price shocks manufactured by OPEC and the disruption of supply chains with the COVID pandemic (pp. 116-117). He adds that even when government spending is greater than tax obligations, prices will not rise if the price paid by the government for the real resources is fixed (p. 62). Thus, he rejects austerity measures to fight "supply-side induced inflation" (p. 116). Moreover, he alludes to the recession of 2009 where despite the increase in debt and quantitative easing, both inflation and interest rates remained low (p. 143). Similarly, Reardon et al. (2018) reject this principle by highlighting that the father of monetarism Milton Friedman admitted it's failure in 2003 (p. 303).

Overall, the critiques of this principle allude to supply side shocks like supply chain disruptions during the pandemic to explain inflation, highlight that inflation did not increase despite massive quantitative easing in the aftermath of the financial crisis, and argue against using this principle to justify austerity measures. In terms of the Disney animation relevant to this principle, a short video clip from Aladdin can be shown to highlight that despite entering Agrabah and showering the city with gold coins the issue of inflation is moot because these gold coins come with real resources (camels, peacocks, elephants, and servants) so that the main issue is that of real resources than money, as upheld by Modern Monetary Theory (MMT).

Figure 9: Aladdin showers gold coins in Agrabah

Credit: DISNEY
Video Clip: https://www.youtube.com/watch?v=mT_8FAMsmCM
Duration: 3:35 minutes

Discussion: The strategy would be to show this short clip as a hook to draw students into a discussion on the idea in MMT that the real constraint in the economy is that of real resources and not money. The connection can be made between the song lyrics that allude to "seventy-five golden camels, fifty-three peacocks, sixty elephants, servants, cooks, and bakers," that is, both resources and labour, and the idea that the main issue is finding resources to build capacity in the economy instead of money, which

is of secondary importance. Moreover, inflation is less of an issue with expanding money supply when corporations hoard cash instead of investing in real projects. Additionally, instead of inflation the issue is of instigating another financial crisis when such cash is used for speculation purposes in the stock market. Thus, through the analogy showcased with this clip, students can be offered a refined understanding that while it is cautious to avoid printing too much money to avoid the fate of Zimbabwe, the real issue is that of resource constraints given supply side shocks so that using austerity measures to fight such shocks would be counterproductive.

2.10 Society faces a short run trade-off between inflation and unemployment

The last principle is on the relationship that arises in the short run when the government prints money, which induces firms to sell more and therefore hire more workers and, in the process, cause higher prices. This alludes to the Phillips Curve that showcases that unemployment is reduced at the cost of higher inflation. Thus, the idea is to tolerate a natural rate of unemployment below which inflationary pressures would drive prices higher. In some sense this last principle brings us full circle as it connects with the first principle on trade-offs at the macro level. However, Goodwin (2014) argues that any inflationary pressure would be contingent on underemployed resources in the economy and that monetary stimulus would facilitate individuals to reduce stress that is part of a hyper competitive economy. Similarly, Wray (2022) rejects the trade-off that government spending to combat unemployment would cause inflation or alternatively the idea that a pool of unemployed workers is required to prevent wages and prices from rising (pp. 113, 115). He mentions that since the financial crisis of 2008, many economists and researchers have concluded that the Phillips Curve is not supported by the data (p. 115). He qualifies that any impact on inflation is contingent on various factors including competition from cheap labour abroad, “strength of labour unions”, and “collusion among producers” (pp. 115, 117).

Overall, the critiques of this principle include the idea that the Phillips Curve has been discredited, that any inflationary pressure of government spending to reduce unemployment is contingent on the presence of underemployed resources and other factors, and the rejection of the idea that some natural rate of unemployment is required to prevent inflation. In terms of the Disney animation relevant to this principle, a short video clip from Christopher Robin can be shown to highlight the idea of putting family first in a stressful economy and to focus on production for the middle class instead of conspicuous consumption, as the former stimulates the economy through consumption and employment. Thus, when families destress by going on a holiday, they contribute to the economy through spending that supports jobs for many underemployed workers, which contrasts with the idea of tolerating some unemployment to prevent inflation.

Figure 10: Christopher Robin puts family first

Credit: DISNEY
Video Clip: https://www.youtube.com/watch?v=_qNSOX3lvHs
Duration: 3:35 minutes

Discussion: The strategy would be to show this short clip as a hook to draw students into a discussion on the idea that spending by the middle class stimulates the economy through real purchases that support jobs, as opposed to the rich that engage in conspicuous consumption or hoard cash for speculative purposes in the stock market. The connection can be made between Christopher Robin making the case for manufacturing luggage for middle class families to go on a holiday that would stimulate the company sales and the idea that the focus should be on the real economy with production and employment. Thus, through the analogy showcased through this clip, students can be offered a critical understanding that the emphasis on reducing unemployment is justified and that the trade-off between unemployment and inflation should be qualified in the presence of underemployed resources.

To recapitulate, the focus in this section was on showcasing the ten principles by drawing the best understanding of both the neoclassical and the heterodox paradigms. Such an approach is in line with Earle, Moran, and Ward-Perkins (2017) who critique that both mainstream and heterodox economists are dismissive of alternative perspectives, which goes against the spirit of pluralism (pp. 115-116). Overall, this section has delineated the best of both neoclassical and heterodox perspectives that are tabulated below to uphold pluralism. Moreover, it has showcased Disney video clips that offer memorable lessons through analogies, which help with recall in the age of information overload, and which elicit student interest due to familiarity and connection with childhood memories.

Table: Neoclassical and Heterodox perspectives with Disney Clips

Neoclassical	Heterodox	Disney Clip
society faces a trade-off between corporate hand outs and spending on education and healthcare	free choice does not always exist	***Beauty and the Beast*** Belle has no free choice in becoming the Beast's prisoner
cost of corporate tax cuts is the foregone healthcare and education spending	free lunch exists with underemployed resources (also with colonialism, slavery, bonded labour)	***Pocahontas*** Governor Ratcliffe and the English settlers appropriate the native land and riches
individuals can ignore sunk costs to start anew through marginal analysis	corporations can take advantage of individuals based on their cognitive abilities and through advertisements	***Snow White and the 7 Dwarfs*** The evil queen manipulates Snow White to eat a poisoned apple
governments need to account for perverse effects of instituting policy	financial incentives can crowd out intrinsic motivations; may not dissuade those with entrenched habits	***Mulan*** Mulan is not incentivized by the Emperor's offer of a seat at the council
free trade allows access to more, greater variety and cheaper goods	powerful actors can exploit weaker parties through trade agreements; workers and the environment are exploited	***The Little Mermaid*** Ariel trading her voice for a pair of legs and being vulnerable to being enslaved is a bad trade
markets avoid the issues of inefficient bureaucracies	markets are dependent on government institutions and rules	***Cinderella*** The invitation to the ball provides a level playing field to every maiden including Cinderella
government intervention is required to address market power and pollution	governments are also essential for safety regulations, anti-discrimination, and equity	***Aladdin*** The Sultan institutes the law that allows commoners to marry into royalty, addressing anti-discrimination and equity
societies should focus on productivity	wages have stagnated despite rise in productivity due to corporate power and weakened labour unions	***Pinocchio*** Stromboli exploits Pinocchio's labour, makes money off his shows, offers poor work conditions
avoid printing too much money to curb inflation	austerity measures to combat supply shock induced inflation are detrimental	***Aladdin*** Aladdin showers gold coins but has labour and resources to back it up
tolerate some unemployment to curb inflation	society should focus on full employment policies; Phillips curve has been discredited	***Christopher Robin*** Christopher Robin tells the board that families going on a holiday would stimulate company sales; the focus is on middle class consumption and jobs

3. Alternative Principles

A critical look at the ten "axiomatic" principles offers a nuanced understanding of the lessons drawn from them. However, it is not enough to simply criticize but to go ahead and offer alternatives (Van Staveren, 2021, pp. 2-3; Raworth, 2017, p. 20). To some extent, the nuanced perspective delineated in the previous section already offers alternative principles for a post-neoliberal economics education. Nonetheless, these alternative principles can be further reviewed and delineated as follows.

Campbell et al. (2019) argue that ECON 101 textbook theory ignores issues of inequality, conditions of work, the quality of life, forced appropriation of resources, power structures that shape market rules, and justifies austerity measures. Therefore, Campbell et al. (2020) offer alternative principles that highlight that capitalism exploits workers and the environment to make profits, that it is inherently unstable and prone to crises, and that the economic goal should be to maximize human potential. Similarly, Goodwin (2014) offers alternative principles that include diminishing marginal utility of income to support redistribution, the trade-off between democracy and wealth concentration, the futility of tax cuts for the rich to generate employment, regulation of the financial sector, and the idea that the real-world economy is not based on Newtonian physics.

In delineating alternative principles in their anti-textbook, Hill and Myatt (2021) highlight that the micro principles of rationality and self-equilibrating markets conflict with herd behaviour and government intervention under crises at the macro level. They shift the economics definition from satisfying unlimited wants to social provisioning, highlight the free lunch obtained from unemployed resources, underscore community disruption caused by free trade, and emphasize the creation of wants through advertisements. They also highlight the detrimental effect of inequality on social cohesion and upward social mobility, emphasize the use of heuristics in decision making, and showcase that financial incentives can crowd out altruistic behaviour. Similarly, Komlos (2021a) mentions relative income instead of absolute income as the basis of well-being, opportunistic behaviour (deception or manipulation) through fine print in contracts, discrimination that prevents a level playing field, predatory advertisements that take advantage of consumers and shape their wants, and government regulation that is needed for consumer protection.

Earle, Moran, and Ward-Perkins (2017) argue for pluralist perspectives in economics education that recognize satisficing, degrowth, power imbalance, and address real world issues including financial crises, climate change, and inequality. Similarly, Skidelsky (2020) offers a vision of an economics that would limit unrealistic assumptions of rational agents, equilibrium, and competitive markets, curb the neoliberal agenda of free trade, liberalization, and smaller governments, shift the focus from efficiency and growth to sustainability and degrowth, and support pluralist perspectives to address issues including automation, inequality, and climate change. Finally, in a collection of fifteen essays Fullbrook and Morgan (2021) reiterate that multiples voices in the heterodox economics tradition converge to a common set of ideas on post neoliberal economics. This includes recognizing power relations, caring communities, tipping points, governments as spenders of last resort, and interdependent satisficing on essential needs. This also includes recognizing the ecological limits to growth, focusing on social provisioning, and emphasizing degrowth instead of technical fixes to address the existential threat of climate change.

Overall, the alternative principles can be crystallized as follows.

1. Economics should be defined for social provisioning not utility maximization.

2. Individuals and firms make decisions based on heuristics and rules of thumb.

3. Financial incentives crowd out intrinsic motivations.

4. Corporations take advantage of individuals based on cognitive abilities, advertisements, and fine print in contracts

5. Relative not absolute income is the basis of well-being.

6. Tipping points and positive feedback loops prevent equilibrium.

7. A free lunch exists with underemployed resources and labour.

8. Capitalism is inherently unstable due to herding behaviour and recurring financial crises.

9. Addressing climate change requires degrowth and respecting ecological limits instead of technical fixes.

10. Inequality harms democracy, social cohesion, and upward social mobility.

11. Diminishing marginal utility of income justifies income and wealth redistribution.

12. Government has an essential role in addressing discrimination, work conditions, and effective functioning of markets.

13. Real resource constraints matter, as money and finances are not an issue.

4. Student Feedback

In a newly designed topics course titled ECON 357: Humanistic Economics, which has a pre-requisite of ECON 101, I engaged students through Disney video clips on the ten principles. I based the course on Komlos (2019) and assigned several books for a book review project to introduce them to post-neoliberal economics. Generally, I found the students quite receptive, and it was as if they were waiting for such a course in their otherwise standard economics education program. Student rating has been incredibly positive compared to the standard intermediate microeconomics courses that I taught in the Fall 2022 semester (4.9/5 compared to 4.7/5 and 4.4/5). The response rate to the end of term teaching evaluation was the highest among the three classes that I taught (50% versus 29.6% and 24%). Additionally, the student feedback comments were quite detailed compared to the regular courses. A snapshot of these comments is as follows (emphasis mine), which support the approach to offer pluralist perspectives on economics principles through Disney video clips.

> "I always look forward to coming to class, the class engagement and discussions (the class culture) ... I'm never bored, I always feel encouraged to engage in discussion, ... He is not a robot regurgitating theory but a human being like the rest of us, caught up in the economy like the rest of us, ... He can be both entertaining and serious and knows how to use this to educate students, **one moment we'll be watching a Disney clip of the little mermaid and the next we're talking about contracted intergenerational indentured servitude in India**."

"Cutting some of that information down so we could get into macro more could be beneficial. **The videos were nice though and helped me to lock ideas in**. Also someone give this man a better computer and projector screen that stays down."

"Easily one of the best courses I've taken … in the many years I've been taking classes here off and on. The course load I felt had lots of material in it, but **I could focus on core principles that meant something to me**. I really enjoyed the book I was assigned to read. So much better than the typically expensive and dry textbooks we sometimes are required to buy and read. … I feel like I actually came away with some real–world knowledge that I'll legitimately use. Awesome course, the best."

"This was the best course I've ever taken in economics, and it has reinvigorated my interest as a field of study, and I really can't emphasize either of those points enough. All the questions and all the grievances that build up in any reasonable person taking micro and macro theory classes are finally answered, I finally got to learn real economics and about the real economy (instead of abstractions of abstractions of abstractions (seriously think about the concept of units of utility on a utility curve)). What's sad is that this class is an exception, a rare glimpse of the truth. What this class highlights is the need for the improvement of other economics courses, with stale theory and sedimented philosophical assumptions that most economists (even top ones like Mankiw) don't engage with. As you can tell I believe the content of this class is of vital societal importance or at least a step in the right direction for teaching economics. So, the great strength of this course is its contents. **Also, I liked going over the 10 principles of economics covered in econ 101 and giving them nuance, keep this**."

"It always felt like we were engaging with the material instead of just choking it down to spit it out on the next quiz. It was also cool to be able to challenge different opinions and raise different points without incurring the displeasure of the professor, …"

"I liked that you moved much of the pressure away from our written assessments to a personal application and understanding of the principles, by asking us directly what we learned and talked about in the class before, and to think critically about real–world principles and how they relate to us."

5. Conclusion

The objective in this paper was to revisit the principles of economics, as many ECON 101 students neither major in economics nor take higher level economics courses. They are therefore left with the dogma of market fundamentalism or economism. While some instructors would like to completely overhaul the way ECON 101 is taught, there remains the issue of large-scale inertia, as many instructors have invested significant human capital in the neoclassical paradigm. There is also a concern that both mainstream and heterodox economists are dismissive of alternative perspectives. Therefore, the focus in this paper has been on illustrating both neoclassical and heterodox perspectives in the spirit of pluralism. Additionally, going beyond the chalk and talk method of instruction, the focus is on Disney video clips, which elicit student interest due to familiarity and connection with childhood memories, generate rich class discussions, and help with recall in the age of information overload. Moreover, the instructor preparation time, class time, and technological requirement of such an approach are minimal but the benefit is high in terms of eliciting student interest, engendering class discussion, and creating

memorable lessons. Student feedback has been incredibly positive. Overall, with the focus on pluralist perspectives and Disney clips, the twin objectives of upholding nuance and retaining student interest are achieved.

References

Acchiardo, C.J., Al-Bahrani, A., Holder. K., and Mateer, D. (2017) "Through the Lens of Life: Teaching Principles of Economics with Humans of New York," *Journal of Economics and Finance Education,* 16(2): 19-30.

Al-Bahrani, A., Holder, K., Patel, D., and Wooten, J. (2016) "Art of Econ: Incorporating the Arts through Active Learning Assignments in Principles Courses," *Journal of Economics and Finance Education,* 15(2): 1-16.

Becker, W., and Watts, M. (1996) "Chalk and Talk: A National Survey on Teaching Undergraduate Economics," *American Economic Review,* 86(2): 448-453.

Bowles, S. and Carlin, W. (2020) 'What students learn in Economics 101: Time for a change', *Journal of Economic Literature*, 58(1), pp. 176-214.

Campbell, A., Davis, A., Fields, D., Quick, P., Ragusett, J., and Schneider, G. (2019) "The unreal basis of neoclassical economics," *American Review of Political Economy*, 13(1).

Campbell, A., Davis, A., Fields, D., Quick, P., Ragusett, J., and Schneider, G. (2020) 'The principles of radical political economics', *American Review of Political Economy*, 14(1).

Colander, D. (2015) "Why economics textbooks should, but don't, and won't, change," *European Journal of Economics and Economic Policies: Intervention*, 12(2): 229-235.

Colander, D. (2016) 'Tools, not rules: Are we teaching the wrong principles of economics in the introductory course?' *Eastern Economic Journal*, 42(2), pp. 163-168.

Dubner, S.J. (2012) "The Cobra Effect", Episode 96: Podcast with Vikas Mehrotra, Freakonomics, October 11, https://freakonomics.com/podcast/the-cobra-effect-2/ (Accessed: December 24, 2022)

Earle, J., Moran, C., and Ward-Perkins, Z. (2017) *The Econocracy*, Manchester University Press, Manchester.

Fullbrook, E. and Morgan, J. (2021) *Post Neo-Liberal Economics*, World Economics Association Books, Bristol.

George, P. (2022) "The mathematics of profit maximization is incorrect," *Real World Economics Review*, 99: 143-150.

Goodwin, M. (2014) 'What is our child learning? Or, Greg Mankiw and the terrible, horrible, no-good, very bad textbook (now with single-page goodness!)', Economix, November 19, https://economixcomix.com/2014/11/19/what-is-our-children-learning-or-greg-mankiw-and-the-terrible-horrible-no-good-very-bad-textbook/ (Accessed: December 22, 2022)

Hill, R. (2022) "Incorporating Critical Perspectives in Teaching Micro Principles: a Nonstandard Text or Supplementary Critical Commentaries?" Heterodox Economics and Teaching Economic Principles, 56th Canadian Economics Association Conference.

Hill, R. and Myatt, T. (2021) *The Microeconomics Anti-Textbook*, 2nd edition, Bloomsbury, London

Jahangir, J. (2021) Amplifying Heterodox Economics with Video Clips, *Challenge*, 64(5-6): 433-452.

Jahangir, J. (2022) John Komlos and the Seven Dwarfs, *Real-World Economics Review,* 101(15): 65-75.

Knudsen, K. and Duncan, T.K. (2018) 'Using Disney as a pedagogical approach', SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3279043 (Accessed: December 22, 2022)

Komlos, J. (2019) *Foundations of Real-World Economics: What every Economics student needs to know*, 2nd edition, Routledge, New York.

Komlos, J. (2021a) "How to change Economics 101," *Challenge,* 64(3): 182-219.

Komlos, J. (2021b) "Covert Racism in Economics," *FinanzArchiv,* 77(1): 83-115.

Komlos, J. (2021c) 'Humanistic economics, a new paradigm for the 21st century', *Real-World Economics Review*, 96: 184-202.

Lee, F.S. (2010) "A heterodox teaching of neoclassical microeconomic theory," *International Journal of Pluralism and Economics Education*, 1(3): 203-235.

Luccasen, R. A., Hammock, M., and Thomas, M.K. (2011) "Teaching Macroeconomic Principles Using Animated Cartoons," *The American Economist,* 56(1): 38-47.

Mandzik, A. (2022) "Once upon an economics course: using fairy tales to teach economics," *Journal of Economics Teaching*, 7(3): 147-161.

Mankiw, N.G., Kneebone, R.D., and McKenzie, K.J. (2020) *Principles of Microeconomics*, 8th Canadian edition, Nelson, Canada.

Raworth, K. (2017) *Doughnut Economics,* Chelsea Green Publishing, Vermont.

Reardon, J., Madi, M.A., and Cato, M.S. (2018) *Introducing a New Economics*, Pluto Press, London.

Skidelsky, R. (2020) *What's Wrong with Economics*, Yale University Press, New Haven.

Van Staveren, I. (2021) *Alternative Ideas from 10 (almost) forgotten economists*, Palgrave Macmillan, Cham.

Wray, R. (2022) *Making Money Work for Us*, Polity, Cambridge.

Author contact: junaid@ualberta.ca

SUGGESTED CITATION:
Junaid Jahangir, "Revisiting the Principles of Economics through Disney", *real-world economics review*, issue no. 103, 31 March 2023, pp. 89-110, http://www.paecon.net/PAEReview/issue103/Jahangir

You may post and read comments on this paper at http://rwer.wordpress.com/comments-on-rwer-issue-no-103/

real-world economics review, issue no. 103
subscribe for free

On the Employer-Employee Relationship[1]

David Ellerman [University of Ljubljana, Slovenia]

Abstract

Much heterodox economics criticizes neoclassical microeconomics as being unrealistic. Yet, as Frank Knight pointed out, it was not meant to be realistic but to posit a normative ideal, the perfectly competitive market economy. Hence serious criticism needs to focus on that competitive ideal, not on its imperfect realization in the real world. Which institutions are at fault in the competitive ideal? This paper focuses on the institution of renting, hiring, employing, or leasing people. Much contemporary progressive narrative criticizes the employment relation in terms of unequal bargaining power and seeks remedies in terms of more equal bargaining power and redistributions of income. This paper analyzes employment in terms of the renting of persons and revives the inalienable rights critique of even fully voluntary contracts to sell or rent oneself out. The implication is the abolition of the employment contract for the renting of persons in favor of the system of workplace democracy where the legal members of a firm are the people working in it.

Keywords: neoclassical microeconomics; the competitive ideal; employment relation; human rentals; inalienable rights; workplace democracy

1. Introduction

Heterodox or non-neoclassical microeconomics is in a state of disarray. The scientific prestige of neoclassical theory is based on the unified standard microeconomic theory. The main criticism of neoclassical microeconomics is that it is not realistic; it does not describe the "real world." But this is fully agreed by neoclassical theorists. The most philosophically and methodologically sophisticated is arguably Frank Knight who clearly made the point that the theory of a competitive market economy is not intended to be descriptive.

> Economic theory is not a descriptive, or an explanatory, science of reality. Within wide limits, it can be said that historical changes do not affect economic

[1] This essay is a summary of the arguments in my two new books published by SpringerNature: *Neo-Abolitionism: The Case for Abolishing Human Rentals in Favor of Workplace Democracy*, and the follow-up book applying the arguments to neoclassical microeconomic theory: *Putting Jurisprudence Back into Economics: What is Really Wrong in Today's Neoclassical Theory*.

> theory at all. It deals with ideal concepts which are probably as universal for rational thought as those of ordinary geometry.[2]
>
> The fact that description of ideal behaviour in part explains actual behaviour operates as a source of confusion; the notion that economics is a science explanatory of actual behaviour is the most important single confusion in the methodology of the science.[3]

The role of competitive microeconomic theory is not as a descriptive theory, but as a regulative ideal to be approached as closely as possible in practice. Those who base their criticism of competitive microeconomics on it not being realistic only invite the sophisticated micro-theorist to say: "Yes, we agree, so help us make the real world economy more like the competitive ideal." The point is simple; serious criticism of neoclassical microeconomics needs to attack the competitive ideal itself, not just its divergence from the real world.

The next question is the point of attack: private property, market economy, employment relation, or what? The point of attack argued for here starts with the last major abolition in the economy, the abolition of slavery. The abolitionist case was not a critique of private property in general but of one type of property in persons, i.e., the ownership of their lifetime of services.[4] The abolitionist case was not a critique of markets in general but of the slavery market—where abolition applied not only to involuntary slavery but also to voluntary contracts for slavery or lifetime servitude. The abolitionist case was not a critique of all legal relationships but of the master-slave relation (whether established involuntarily or voluntarily).

The current economic system was arrived at by abolishing the ownership of workers (meaning the ownership of all their services) in favor of the voluntary renting, hiring, employing, or leasing of workers for some time period. That institution is totally taken for granted in neoclassical microeconomics. This paper summarizes the neo-abolitionist case for the abolition of not only the ownership of workers but also the renting of people in favor of workplace democracy.

2. The difference one word makes

In teaching Economics 101, I would describe a slavery economy, e.g., the Antebellum South, as a private property market economy where the workers were owned by their employer. Then after noting that just as one could own a car or rent a car, I asked if anyone knew of an economy where the workers were rented instead of owned? One quick answer was "feudalism", but I would explain that the serfs were attached to the land and thus part of the manor. They were a superior class to slaves (or "servi" in Latin) since the serfs were not chattel that could be 'sold down the river' Indeed, a retronym was needed to separate the serfs and "servi" so the slaves in the European Middle Ages were renamed after their most common ethnic origin (Slavs). But the original question is: does anyone know of an economy based on renting people? After a

[2] Knight, Frank H. 1969. *The Ethics of Competition and Other Essays*. Freeport NY: Books for Libraries Press, p. 277.

[3] Ibid., p. 279.

[4] The slave-owners and their apologists made the point that slavery does not involve the ownership of a person's soul ("Souls don't chop cotton; labor does"); it was a means to acquire a secure source of labor.

pause, an African-American student might point out that during slack times, field slaves were rented out as stevedores on the docks or to work in military construction projects.[5]

If I had asked instead if anyone know of an economy where workers were *hired* or *employed*, then the answer would be a 'no-brainer.' Just one word, asking about "rented" people instead of "hired" people makes a huge difference in the consciousness of people living in today's economic system. Yet technically, the employment relation is the renting of people, the purchase of a person's services (e.g., a person-day)—just as renting a car for a day is the purchase of a car-day or the renting of an apartment for a month is the purchase of an apartment-month of services. In the US, we say people are hired and cars are rented, but in the UK, a rental car is called a "hire car."

Some Economics 101 textbook writers are forthright enough to point out these simple facts (although almost all are not). The greatest of the neoclassical economists, Paul Samuelson, was quite clear on the point.

> Since slavery was abolished, human earning power is forbidden by law to be capitalized. A man is not even free to sell himself: he must *rent* himself at a wage.[6]

Other textbook writers could be equally forthright.

> The commodity that is traded in the labor market is labor services, or hours of labor. The corresponding price is the wage per hour. We can think of the wage per hour as the price at which the firm rents the services of a worker, or the rental rate for labor. We do not have asset prices in the labor market because workers cannot be bought or sold in modern societies; they can only be rented. (In a society with slavery, the asset price would be the price of a slave.)[7]

One of Samuelson's teachers, Frank Knight, used the synonym of leasing.

> [I]n a free society the larger part of the productive capacity employed (as matters stand today in a typical Western nation) consists of the services of human beings themselves, who are not bought and sold but only, as it were, leased.[8]

[5] Hulse, Thomas. 2010. "Military Slave Rentals, the Construction of Army Fortifications, and the Navy Yard in Pensacola, Florida, 1824-1863." *The Florida Historical Quarterly* 88 (4 Spring): 497–539.

[6] Samuelson, Paul A. 1976. *Economics*. New York: McGraw-Hill, p. 52 (his italics).

[7] Fischer, Stanley, Rudiger Dornbusch, and Richard Schmalensee. 1988. *Economics*. New York: McGraw-Hill, p. 323.

[8] Knight, Frank. 1936. "The Quantity of Capital and the Rate of Interest: I." *Journal of Political Economy* 44 (4): 433–63. p. 438.

3. Owned workers versus rented workers

The bald comparison of owned workers under slavery and today's rented workers misses the point that the employment relation is voluntary—which was hardly the case for historical slavery. And contrary to much progressive narrative by philosophers and legal theorists,[9] the human rental contract is indeed voluntary by any juridical standard and even compares favorably to the unbargained contracts of adhesion that consumers face at supermarkets or department stores.

Thus, the real point of comparison is between a system of *voluntary* contractual lifetime or long-term servitude with today's system of voluntary limited-term human rental contracts. Conventional classical liberalism, as opposed to a much deeper tradition of inalienable rights in democratic or Enlightenment classical liberalism (seemingly little known among today's progressives), has no principled differentiation between long-term (e.g., for a working lifetime) and short-term voluntary labor contracts—as long as they are voluntary. This is as true for the founders as for today's descendants of Adam Smith. As Frank Knight points out: "Interestingly enough, the political and legal theory had been stated in a series of classics, well in advance of the formulation of the economic theory by Smith. The leading names are, of course, Locke, Montesquieu, and Blackstone."[10] Yet all three of those founders of conventional classical liberalism condoned a voluntary contract of lifetime servitude (perhaps using some euphemisms). Here are the relevant quotes.

> For, if once *Compact* enter between them, and make an agreement for a limited Power on the one side, and Obedience on the other, the State of War and *Slavery* ceases, as long as the Compact endures.... I confess, we find among the *Jews*, as well as other Nations, that Men did sell themselves; but, 'tis plain, this was only to *Drudgery, not to Slavery*.[11]

> This is the true and rational origin of that mild law of slavery which obtains in some countries; and mild it ought to be, as founded on the free choice a man makes of a master, for his own benefit; which forms a mutual convention between two parties.[12]

> Yet, with regard to any right which the master may have lawfully acquired to the perpetual service of John or Thomas, this will remain exactly in the same state as before: for this is no more than the same state of subjection for life, which every apprentice submits to for the space of seven years, or sometimes for a longer term.[13]

[9] Mishel, Lawrence, ed. 2022. "Not So Free to Contract: The Law, Philosophy, and Economics of Unequal Workplace Power." *Journal of Law and Political Economy* 3 (1).

[10] Knight, Frank. 1947. *Freedom and Reform*. New York: Harper & Row., p. 27, fn. 4.

[11] Locke, John. 1690. *Second Treatise on Government*, § 24.

[12] Montesquieu, Count. 1748. *Spirit of the Laws*, Vol. I, Bk. XV, Chap. V.

[13] Blackstone, William. 1765. *Commentaries on the Laws of England*, Section on "Master and Servant."

These are the aspects of classical liberalism that today's neoclassical (and Austrian) economists simply do not address. They are unable or unwilling to address what might be called "Philmore's Challenge."

> Contractual slavery and constitutional non-democratic government are, respectively, the individual and social extensions of the employer-employee contract. Any thorough and decisive critique of voluntary slavery or constitutional non-democratic government would carry over to the employment contract—which is the voluntary contractual basis for the free market free enterprise system.[14]

Thus, it should come as little surprise when today's classical liberals or libertarians advocate that people should be able to voluntary alienate their legal self-governance rights not only in the workplace (as in the employment contract) but in municipal or state governments as in charter or startup cities or in today's Dubai.

> The comparable question about an individual is whether a free system will allow him to sell himself into slavery. I believe that it would.[15]

But it may be a little more surprising to find the same assumption in neoclassical economics. The crown jewel in neoclassical economics is the theorem that a competitive equilibrium is allocatively efficient (or Pareto optimal). This theorem requires the assumption of complete futures markets in all commodities. If there was a limitation on the time span of human rental contracts, say at ten years of service, then there might be willing buyers and sellers of labor services dated eleven years in the future so outlawing that transaction would preclude allocative efficiency. The fact that the crown jewel theorem of neoclassical microeconomics requires allowing what Blackstone called "perpetual service" contracts is not mentioned, to the author's knowledge, in any textbook. But the econometrician, Carl Christ, pointed it out in no less a forum than Congressional testimony.

> Now it is time to state the conditions under which private property and free contract will lead to an optimal allocation of resources.... The institution of private property and free contract as we know it is modified to permit individuals to sell or mortgage their persons in return for present and/or future benefits.[16]

According to Frank Knight, the present limitation on long-term or lifetime labor contracts is "logically not a part of the property system." Indeed, it is "one of the defects of our civilization…".[17]

> If laborers were not guaranteed the "inalienable right" of freedom, that is, if they could make enforceable time contracts for work and thus capitalize their labor

[14] Philmore, J. 1982. "The Libertarian Case for Slavery: A Note on Nozick." *Philosophical Forum* XIV (Fall): 43–58, p. 55.

[15] Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books, p. 331.

[16] Christ, Carl F. 1975. "The Competitive Market and Optimal Allocative Efficiency." In *Competing Philosophies in American Political Economics*, ed. John Elliott and John Cownie, 332–38. Pacific Palisades, CA: Goodyear. pp. 337-8.

[17] Knight, Frank. 1965 (1921). *Risk, Uncertainty and Profit*. New York: Harper Torchbooks, p. 350, fn. 1.

> power they would in an economic sense be more secure—in the sense in which the slave has security.[18]

The point is that conventional classical liberalism and even contemporary economic theory have no principled differentiation between long-term human rental contracts forbidden by inalienable rights and the short-term human rental contracts of today's economy where such inalienable rights supposedly do not apply.

4. Persons and things in production

Even since the marginalist revolution at the end of the nineteenth century, economics has been dominated by a certain "picture" of production where "input services cooperate together to produce the product" and then attention is to be focused on the "division or distribution of the product." What could be wrong with that picture? Clearly all the inputs must productively contribute to the product, and the revenue from the product must eventually be distributed to the owners of those inputs.

What's wrong with that picture is that it does not distinguish between persons and things. It does not differentiate between the responsible actions of persons and the causally efficacious services of things. Knight loved the word "productive" since it blurred the difference between responsible actions and the efficacious services of things.

Of course, in their non-professional lives, neoclassical economists know the difference between responsible actions and the mechanical services of things. When called in for jury duty for the trial of a person accused of committing murder with a gun, the economist will presumably not ask the judge why the gun is not on trial. The economist knows that responsibility is imputed back through the "instruments" to their human user. But search through the entire literature of neoclassical (or heterodox?) economics to try to find such a simple aspect of jurisprudence recognized in the context of normal production. The author has only been able to find a single example of an economist, indeed, the juridically-trained Austrian economist, Friedrich von Wieser, who stated the simple facts.

> The judge ... who, in his narrowly-defined task, is only concerned with the legal imputation, confines himself to the discovery of the legally responsible factor,—that person, in fact, who is threatened with the legal punishment. On him will rightly be laid the whole burden of the consequences, although he could never by himself alone—without instruments and all the other conditions—have committed the crime. The imputation takes for granted physical causality. …
>
> If it is the moral imputation that is in question, then certainly no one but the labourer could be named. Land and capital have no merit that they bring forth fruit; they are dead tools in the hand of man; and the man is responsible for the use he makes of them.[19]

[18] Knight, Frank. 1956. *On the History and Method of Economics*. Chicago: Phoenix Books, p. 93, fn. 6.

[19] Wieser, Friedrich von. 1930. *Natural Value*. New York: G.E. Stechert, pp. 76-9.

That was the juridically-trained von Wieser speaking. But for the *economist*, von Wieser, the implication was that economics apologetics *obviously* needs a different notion of responsibility than the usual legal or moral responsibility based on the difference between persons and things. Economics needs the notion of the "economically responsible factors" so that "In the division of the return from production, we have to deal similarly ... with an imputation, – save that it is from the economic, not the judicial point of view."[20] Instead of seeing the usual legal and moral notions of responsibility and imputation applying to the system of property and contracts that underlies a market economy, the economics profession 'needs' to devise a different 'picture' of the "division of the return from production" that ignores the difference between persons and things and treats them all as "economically responsible factors."

But how else can economics account for the "division of the product"? The answer is to look at the actual rather than metaphorical "division of the product" in terms of property assets and liabilities. There is no actual division to the ownership of the product; it is *all* owned by the employer or employing corporation. And the production of the product is not the only property-related result of production. There is also the using up of the inputs which creates the liability to pay for those inputs—which is typically paid off by the employer purchasing the necessary inputs beforehand (except for labor). In fact, the assets representing the produced product and the liabilities representing the used-up inputs are both *correctly* represented by the respective positive and negative entries in the modern economic notion of "input-output vector"[21] or production vector. And the point is that, in fact, one party, typically the employer (natural person or corporation), gets that whole vector—which we might, for historical reasons, call the "whole product."[22]

The legal party that gets or appropriates the whole product (input liabilities plus output assets) of production is typically called the "firm." Given today's obscene distribution of income and resulting wealth, progressive economists focus on unequal bargaining power and the "question of distribution" which is treated in neoclassical microeconomics as a metaphorical "division of the product" according to the metaphorical notion of "economically responsible factors." But insofar as the actual legal facts are allowed in the 'science of economics,' the people rented into a firm own *zero* percent of the output-assets and owe *zero* percent of the input-liabilities. Try to find a single economics book that states those simple facts. What economists apparently cannot say was clearly said by an economic sociologist over a century ago.

> There is much theoretic discussion to the "right of labor to the whole product" and much querying as to how much of the product belongs to the laborer. These questions never bother the manufacturer or his employee. They both know that, in actual fact, all of the product belongs to the capitalist, and none to the laborer. The latter has sold his labor, and has a right to the stipulated payment therefor. His claims stop there. He has no more ground for assuming

[20] Ibid., p. 76.

[21] Quirk, James, and Rubin Saposnik. 1968. *Introduction to General Equilibrium Theory and Welfare Economics*. New York: McGraw-Hill, p. 27.

[22] Menger, Anton. 1899. *The Right to the Whole Produce of Labour: The Origin and Development of the Theory of Labour's Claim to the Whole Product of Industry*. Translated by M.E. Tanner. London: Macmillan and Co.

> a part ownership in the product than has the man who sold the raw materials, or the land on which the factory stands.[23]

The wages are not a "share of the product" but are one of the input-liabilities owed and paid off by the employer. The fundamental question is not about getting a bigger piece of the pie any more than the question about slavery was whether the real income of the slaves (food, clothing, or shelter) was less than their marginal productivity (e.g., the pathetic debate[24] that followed the Fogel and Engerman's econometric analysis of slavery).[25]

5. Who is to be the firm in the first place?

By focusing on distribution and redistribution of the inequality symptoms of the human rental system,[26] economists miss the prior and fundamental predistributive question of who is to be the firm in the first place;

- Capital (the owners of capital renting people to undertake production),
- Labor (people renting or owning the capital needed for production), or
- the State renting or owning both capital and labor as in socialism/communism.

In today's free market economy, labor may hire capital just as capital may hire labor, so what is the problem? It is part of the operation of a private property market economy that one legal party cannot pay off the input-liabilities and then another party get the first ownership of the produced outputs. One party has to appropriate that whole production vector or whole product, so the question is who should be that party, the firm, in the first place. The question is *not* whether or not "the labor-power… were paid for at its full value."[27]

In the legal system, the assignment of liabilities for property damages is governed by the principle of assigning legal responsibility according to factual responsibility, the juridical principle of imputation. The ideological role of marginal productivity theory is to try to show that competitive equilibrium imputes a metaphorical "share of the product" to each factor according to its metaphorical "economic responsibility."

> The basic postulate on which the argument rests is the ethical proposition that an individual deserves what is produced by the resources he owns.[28]

[23] Fairchild, Henry Pratt. 1916. *Outline of Applied Sociology*. New York: Macmillan, pp. 65-66

[24] For instance, David, Paul A., Herbert G. Gutman, Richard Sutch, Peter Temin, and Gavin Wright. 1976. *Reckoning with Slavery*. New York: Oxford University Press.

[25] Fogel, Robert W., and Stanley L. Engerman. 1974. *Time on the Cross.* Boston: Little, Brown and Company.

[26] Blanchard, Olivier, and Dani Rodrik, eds. 2021. *Combating Inequality: Rethinking Government's Role*. Cambridge MA: MIT Press.

[27] Marx, Karl 1977 (1867). *Capital (Volume I).* B. Fowkes Trans., New York: Vintage Books, p. 357 fn. or Chap. 10, sec. 3.

[28] Friedman, Milton. 1976. *Price Theory*. Chicago: Aldine, p. 199.

> The analysis [of market competition] shows how, under the conditions necessary for its existence, this organization achieves … *justice* in the distribution of the total product, … justice by the principle of equality in relations of reciprocity, giving each the product contributed to the total by its own performance ("what a man soweth that shall he also reap").[29]

These arguments pay homage to the imputation principle; the only problem is that input-suppliers only own a *metaphorical* share of the product, and the imputation is according to a *metaphorical* notion of 'economic' responsibility.

What is the result if we apply the actual non-metaphorical notion of legal or moral imputation to the actual property rights and liabilities generated in production? As von Wieser pointed out, "no one but the labourer could be named. Land and capital have no merit that they bring forth fruit; they are dead tools in the hand of man; and the man is responsible for the use he makes of them." In other words, it is the people who work in an economic enterprise, employees and managers, working employers and 'hired hands,' who perform the deliberate human actions that use up input services of things such as land and capital in the process of producing the output-assets. Hence the actual legal or moral imputation principle assigns those input-liabilities and those output-assets to the people working in the firm. In legal terms, they should be the legal members of the firm as a corporate entity.

Such a firm is called a "democratic firm" or "workplace democracy" and is exemplified in the Mondragon system of worker cooperatives.[30] The application of democratic theory to the workplace[31] was well stated by Justice Louis Brandeis.

> The civilized world today believes that in the industrial world self-government is impossible; that we must adhere to the system which we have known as the monarchical system, the system of master and servant, or, as now more politely called, employer and employee. It rests with this century and perhaps with America to prove that as we have in the political world shown what self-government can do, we are to pursue the same lines in the industrial world.[32]

The legal or moral notion of imputation gives the same answer to the question of who should be the members of the firm; the people working in the firm. This was pointed out by the UK's Tory MP known as the "Minister of Thought", Lord Eustace Percy.

> Here is the most urgent challenge to political invention ever offered to the jurist and the statesman. The human association which in fact produces and distributes wealth, the association of workmen, managers, technicians and directors, is not an association recognised by the law. The association which the law does recognise—the association of shareholders, creditors and directors—is incapable of production and is not expected by the law to perform

[29] Knight, Frank. 1956. *op. cit.*, p. 292.

[30] Whyte, William Foote, and Kathleen King Whyte. 1991. *Making Mondragon. 2nd revised ed.* Ithaca: ILR Press.

[31] Dahl, Robert A. 1985. *Preface to Economic Democracy*. Berkeley: University of California Press, p. 91.

[32] Brandeis, Louis D. 1934. *The Curse of Bigness*. New York: Viking, p. 35

> these functions. We have to give law to the real association, and to withdraw meaningless privilege from the imaginary one.[33]

This 'outlandish' notion of applying the *actual* principle of legal and moral imputation to the *actual* liabilities and assets created by the responsible action of persons is the modern rendition of the old "labor theory of property." It is the only notion of legitimately initiating (or terminating) property rights outside of exchange or gifts. Thus far from being "based on private property," the current system based on renting persons violates the very standard of legitimation for private property.

6. Back to inalienable rights

There is the dangling question of what is wrong with a free and voluntary contract for the purchase and sale of human labor? Is it primarily a matter of unequal bargaining power between employer and employee? Should we play the parlor game of raising our standards of voluntariness so the labor contract is seen as coercive?

This takes us back to the deeper tradition of democratic classical liberalism in the theory of inalienable rights. The basis for the theory is a claim about the facts of human nature, not a claim about rights or bargaining power. It is a fact of human nature that people cannot by some voluntary act alienate their *factual* responsibility for their deliberate actions. In a contract to rent out one's car or apartment, the rented thing can in fact be turned over to the renter who is responsible for the use they make of the rented *thing*. But one cannot do the same by renting out oneself.

One of the founders of Swedish social democracy, Ernst Wigforss, made the point long ago that labor does not have the (factual) alienability required by a purchase and sale contract.

> There has not been any dearth of attempts to squeeze the labor contract entirely into the shape of an ordinary purchase-and-sale agreement. The worker sells his or her labor power and the employer pays an agreed price. What more could the worker demand, and how could he or she claim a part in the governance of the company? …But, above all, from a labor perspective the invalidity of the particular contract structure lies in its blindness to the fact that the labor power that the worker sells cannot like other commodities be separated from the living worker. This means that control over labor power must include control over the worker himself or herself. Here perhaps we meet the core of the whole modern labor question, and the way the problem is treated, and the perspectives from which it is judged, are what decide the character of the solutions.[34]

The Law fully recognizes this *fact* (about the inalienability of human agency) in the case of employer and employee cooperating together to commit a crime. The servants in work suddenly become the partners in crime.

---

[33] Percy, Lord Eustace. 1944. *The Unknown State: 16th Riddell Memorial Lectures.* London: OUP, p. 38.

[34] Wigforss, Ernst. 1923. *Den Industriella Demokratiens Problem 1*. Stockholm: A.-B. Hasse W. Tullbergs boktryckeri, p. 28 (quote translated by Patrik Witkowsky).

> All who participate in a crime with a guilty intent are liable to punishment. A master and servant who so participate in a crime are liable criminally, not because they are master and servant, but because they jointly carried out a criminal venture and are both criminous.[35]

Yet when employee and employer have 'jointly carried out a *non*-criminal venture', then the employees do not suddenly morph into robots or non-responsible instruments. It's not the facts that change; it is the Law that then recognizes the human rental contract as valid (unlike the case of a hired criminal contract), and the employer appropriates 100% of the input-liabilities and 100% of the output-assets. The law for renting persons, like the old law for owning persons, pretends to legally determine when persons are legally persons and when they are legally rented or owned 'things.' As stated by an Antebellum judge, the slaves in fact

> are rational beings, they are capable of committing crimes; and in reference to acts which are crimes, are regarded as persons. Because they are slaves, they are ... incapable of performing civil acts, and, in reference to all such, they are things, not persons.[36]

It was this sort of mismatch between factual persons being legally treated as things that accounted for the abolition of voluntary contracts of lifetime servitude in addition to abolishing involuntary slavery. And it was this fact about human nature that one's personhood cannot be voluntarily alienated that rendered the lifetime contracts invalid and those rights inalienable. Those facts about the voluntary inalienability of factual personhood are the same for the short-term human rental or employment contract as for the long-term contract for selling labor services.

Author contact: david@ellerman.org

SUGGESTED CITATION:
David Ellerman, "On the Employer-Employee Relationship", *real-world economics review*, issue no. 103, 31 March 2023, pp. 111-121, http://www.paecon.net/PAEReview/issue103/Ellerman

[35] Batt, Francis. 1967. *The Law of Master and Servant. 5th ed*. London: Pitman, p. 612. This is the same lawbook quoted in an earlier edition to describe the role of the servant or employee in: Coase, Ronald H. 1937. "The Nature of the Firm." *Economica* IV (Nov.): 386–405, p. 403.

[36] Catterall, Helen T. 1926. *Judicial Cases Concerning Slavery and the Negro*. Vol. III. Washington, DC: Carnegie Institute, p. 247.

real-world economics review, issue no. 103
subscribe for free

Why is yield-curve inversion such a good predictor of recession?

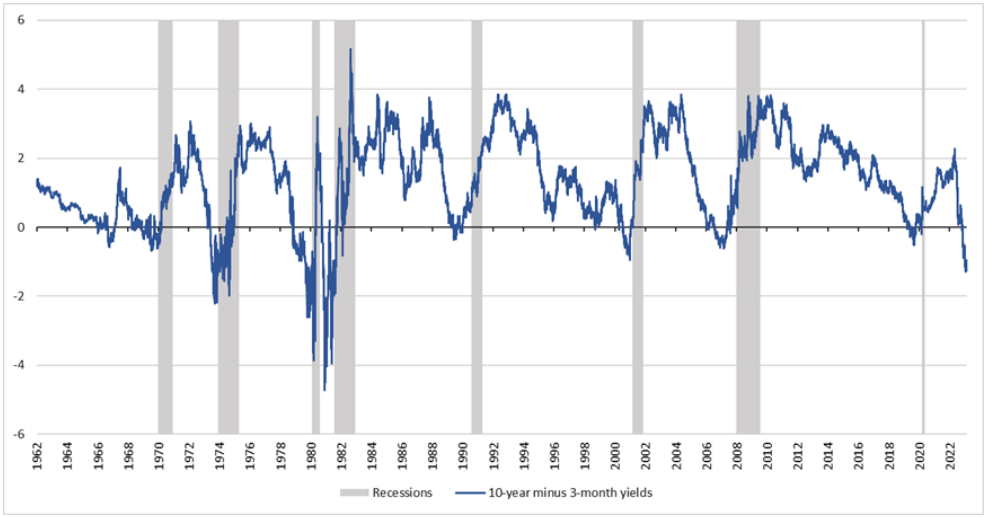
Philip George

Abstract
Yield curve inversion has traditionally been ascribed to investor expectations. This paper shows instead that it is the result of changes in real variables like investment and consumption, and that these changes are set in motion by central bank actions.

1. Introduction

For several decades it has been observed that when the long-term interest rate (typically the yield on 10-year treasury bonds) falls below the short-term interest rate (typically the yield on 3-month treasury bills) a recession follows after a period ranging from a few months to a couple of years.

The figure below depicts the phenomenon. Each of the past eight recessions has been preceded by a yield curve inversion. There is also one "false positive". A yield curve inversion in 1966 was not followed by a recession. It would not be out of place, though, to mention that the credit crunch of 1967 has sometimes been described as a "mini-recession", notably by Milton Friedman. The yield curve inversion of 1966 may not be therefore be so false a positive as it seems.



For plotting the above graph, the daily market yield on US treasury securities at 10-year constant maturity is taken from https://fred.stlouisfed.org/series/DGS10. The daily secondary

market rate on 3-month treasury bills is taken from https://fred.stlouisfed.org/series/DTB3. Since the yield on treasury bills is calculated assuming a year of 360 days, this was adjusted to give the bond equivalent yield. The formula for the conversion is taken from a New York Federal Reserve Bank paper (Estrella, Trubin, 2006) which also usefully summarises the evidence and proffered rationale for yield curve inversion. The dates of recessions are taken from https://fred.stlouisfed.org/series/USRECD.

2. The usual rationale

Yield curve inversion is generally sought to be explained in terms of investor expectations. For example, it could be the result of investor expectations of monetary policy. The central bank raises short-term interest rates to curb inflationary pressures. But investors expect this tightening to reduce inflation in the long term and cause the central bank to ease short-term interest rates in the future. These expectations of central bank monetary policy lead to a reduction in long-term interest rates and cause a flattening or inversion of the yield curve.

Another argument is that expectations of future short-term interest rates depend on expectations of future real demand for credit and future inflation.

> "A rise in short-term interest rates induced by monetary policy could be expected to lead to a future slowdown in real economic activity and demand for credit, putting downward pressure on future real interest rates. At the same time, slowing activity may result in lower expected inflation, increasing the likelihood of a future easing in monetary policy. The expected declines in short-term rates would tend to reduce current long-term rates and flatten the yield curve. Clearly, this scenario is consistent with the observed correlation between the yield curve and recessions." (Estrella, Trubin, 2006)

The trouble with using expectations, an explanatory variable that exists only in people's heads and cannot be measured, is that it can be tortured to fit whatever conclusions one desires. Why does the 10-year yield sometimes rise when the short-term interest rate is raised? Answer: Expectations. Why does the 10-year yield sometimes fall when the short- term interest rate is raised? Answer: Expectations. But an explanation that can account for anything and everything is really no explanation at all.

There is another problem. The inversion of the yield curve is regarded as a result of investor expectations and these turn out to be accurate eight in nine times. But if someone claims to have accurately predicted rain in the coming month eight times out of nine, one would expect him to equip himself with an umbrella to prepare for the rain he predicted. However, in the case of yield curve inversion, which is attributed to accurate investor expectations, it is generally the case that the central bank and government have to step in to supply umbrellas.

One is therefore forced to conclude that yield curve inversion is not to be explained by expectations but by variations in real, measurable variables. That is what the rest of this paper seeks to do.

3. The relationship between investment and long-term interest rates

It is a given in economic theory that investment is a function of interest rates. Specifically, the higher the interest rate the lower the investment and vice versa.

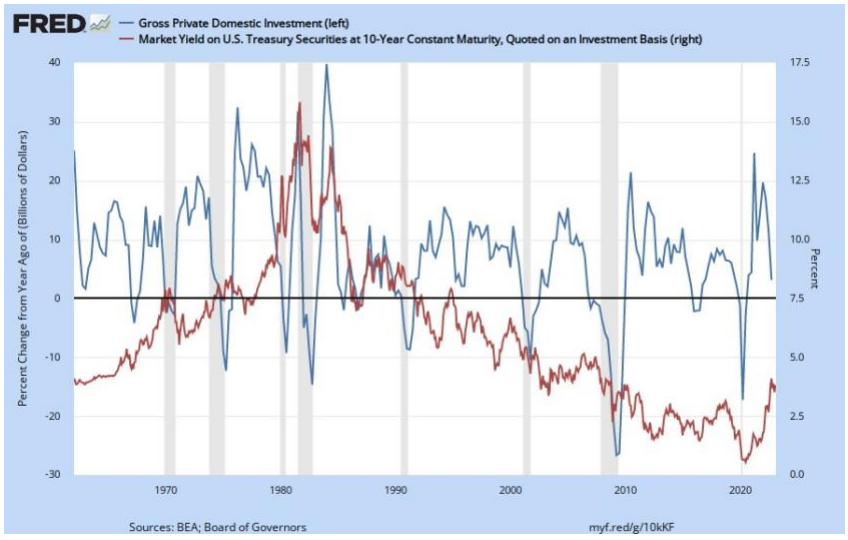
This relationship is one of the equations in J.R. Hicks's famous paper attempting to interpret Keynes's *General Theory* (Hicks, 1937). The assumption of course runs through the *GT* which recommends keeping interest rates low. It is also presumably why central banks cut interest rates each time a recession threatens.

However, Keynes's views are a little more sophisticated than this one-dimensional relationship suggests. His argument runs as follows. A capitalist, while considering whether to invest in a capital asset, estimates the stream of returns flowing from it at various points of time in the future. The marginal efficiency of capital is the discount rate that equates the present value of those returns with the cost of that asset. The capitalist then compares the marginal efficiency of capital with the interest rate. If the marginal efficiency of capital is less than the interest rate then it obviously does not make sense to buy the capital asset. For instance, if he is using his own funds, it would be more profitable to lend them out than invest in the capital asset.

So, it is not the interest rate alone which determines investment. "The schedule of the marginal efficiency of capital is of fundamental importance because it is mainly through this factor (much more than through the rate of interest) that the expectation of the future influences the present." (Keynes, 1936, Chapter 11). Nevertheless, "no one doubts that the investment demand-schedule falls with a rising rate of interest" (Keynes, 1936, Chapter 14).

However, even among those economists who agree with Keynes that effective demand drives the business cycle, not everyone agrees that interest rates determine investment. Michal Kalecki, for instance, "held that the rate of interest is not a significant influence on investment". (Toporowski, 2013)

But what does the empirical data reveal about the connection between interest rates and investment? The graph below shows the YoY percentage change in gross private domestic investment for the US vs the 10-year treasury yield. It shows, for the most part, that investment is high when the interest rate is high and low when the interest rate is low.



This seems to fly in the face of common sense until we realise that we are simply looking at the arrow of causation in the wrong direction. It is not that companies are raising their investments when long-term interest rates are high. It is that long-term interest rates rise when companies increase their investment. This was clear to Kalecki: “Thus changes in the rate of interest are determined by the mechanism of the business cycle rather than determining it.” (Toporowski, 2013).

So, the long-term interest rate is really an indicator of the volume of investment. When capitalists invest more, the long-term rate of interest rises; when they invest less, the long- term rate of interest falls.

4. The effect of changes in short-term interest rates on consumption

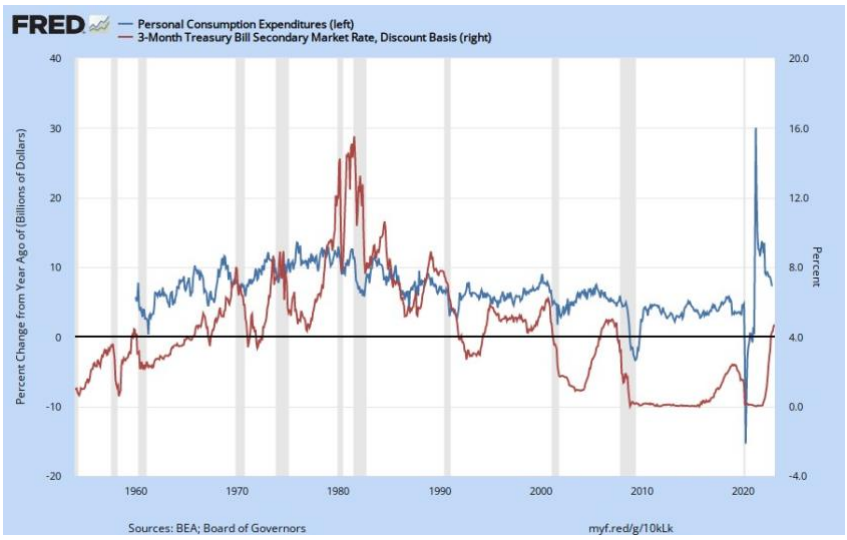
Changes in the short-term interest rate affect the economy primarily due to their effect on leverage and thus on the prices of financial assets, and through them, on personal consumption.

Consider a hedge fund with an equity capital of $5 that borrows $95 and invests the money ($100) in the stock market.

At an interest rate of 0%, when the market rises 1% the fund earns a return of 20% on equity.
When the interest rate is 1% the annual cost of financing is 0.95%. So the market needs to rise nearly 2% to give a return of 20% on equity. If the market remains stationary, the fund suffers a loss of nearly 20% on equity.When the interest rate is 2% the annual cost of financing is 1.9%, so the market needs to rise nearly 3% to give a return of 20% on equity. If the market does not rise, the fund suffers a loss of 38% on equity.

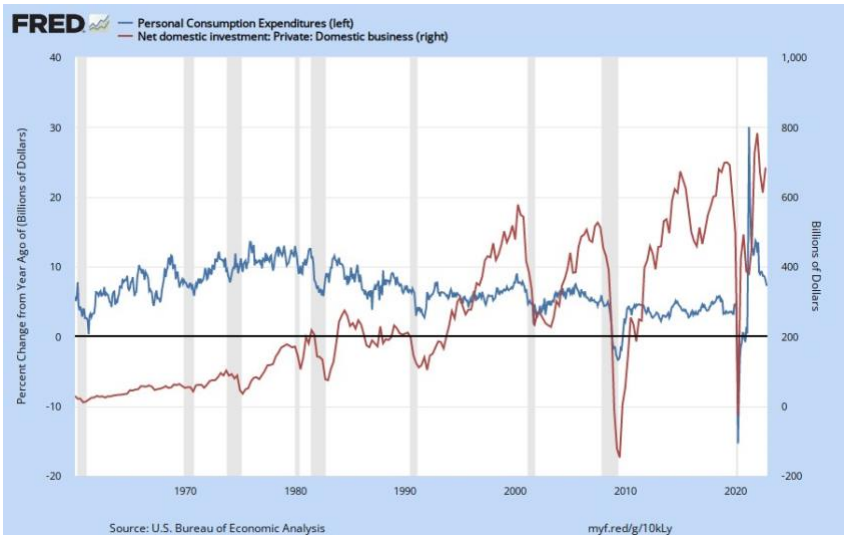
When the interest rate is 5% the annual cost of financing is 4.75%, so the market must rise 5.75% to give a return of 20% on equity. If the market does not rise, almost all the fund’s equity gets wiped out.

When the short-term interest rate rises, inflows into financial asset markets fall, causing a fall in asset prices, and through them, a reduction in personal consumption or, at least at first, a reduction in its growth rate. In the US, home ownership is far wider than equity ownership.
So, the primary effect of an increase in the Fed rate on consumption is through the fall in home prices, the rise in monthly mortgage payments, and the drying up of refinance options.
The figure below shows YoY changes in personal consumption expenditures against the 3-month treasury rate. To cool down a heating economy, the Fed keeps raising the Fed rate with consequent effects on other short-term rates like the 3-month treasury. This brings down personal consumption expenditure, though often with a lag. Fear of an oncoming recession then forces the Fed to reduce interest rates but by then the fall in asset prices becomes uncontrollable and personal consumption expenditure continues to fall.



5. The effect of changes in consumption on private investment

Investment is strongly responsive to changes in personal consumption expenditure. Capitalists invest in anticipation of increases in consumption. If consumption falls considerably net investment may even turn negative, as the figure below (net domesticprivate investment vs YoY change in personal consumption expenditure) shows. Investment is a function of changes in consumption or, more accurately, the rate of change of consumption.



6. Why is yield curve inversion a good predictor of recessions?

The question above can be rephrased as: Why is it that recessions are best predicted not by the short-term rate of interest alone nor by the long-term rate of interest alone but by a combination of the two? In light of what we have said above, the answer is that the long-term rate of interest is linked to the short-term rate of interest. The mechanism too is clear. A rise in the Fed rate raises the cost of leverage and affects markets for financial assets. This causes a

fall in personal consumption, though with a lag. In response capitalists cut back on investment, which causes a fall in the long-term interest rate, and thus an inversion of the yield curve.

At the point where the yield curve inverts, the long-term interest rate is equal to the short- term interest rate. At that point, investment and thus the demand for long-term funds has fallen so far that it is no longer profitable to borrow short-term funds to lend them long-term. Eventually, the continuing fall in investment combines with the continuing fall in the growth rate of personal consumption to cause a fall in GDP, which is of course what a recession is.

From this several policy recommendations can be derived. Reducing short-term rates to near-zero levels each time a recession threatens does nothing to stimulate investment because investment is driven by considerations of profit which in turn is driven by personal consumption expenditure. Ultra-low interest rates mainly help financial market participants and cause asset bubbles. A subsequent rise in the short-term interest rate bursts the bubbles that have been caused by low interest rates and sets off recessions, ranging from mild to severe, and short to long-lasting, depending on how much the accumulated savings of ordinary consumers have been eroded by the bursting of asset bubbles.

7. Conclusion

If yield-curve inversion is a good predictor of recessions it is not because of expectations but due to the effect that short-term interest rate increases have on real variables like consumption and investment and thus on long-term interest rates.

References

Estrella, Arturo and Trubin, Mary R. (2006), "The Yield Curve as a Leading Indicator: Some Practical Issues", *Current Issues in Economics and Finance,* July/August, Federal Reserve Bank of New York, page 3

Hicks, J.R. (1937), "Mr Keynes and the 'Classics'; A Suggested Interpretation",*Econometrica,* Vol 5, No 2 (April), pp 147-159

Keynes, John Maynard (1936), *The General Theory of Employment, Interest and Money*, London, MacMillan

Toporowski, Jan (2013), *Michal Kalecki: An Intellectual Biography, Volume I, Rendezvous in Cambridge, 1899-1939*

Author contact: philip.george@gmail.com

SUGGESTED CITATION:
Philip George, "Why is yield-curve inversion such a good predictor of recession?", *real-world economics review*, issue no. 103, 31 March 2023, pp. 122-127, http://www.paecon.net/PAEReview/issue103/George

You may post and read comments on this paper at http://rwer.wordpress.com/comments-on-rwer-issue-no-103/

Book Review

Thomas Picketty

A Brief History of Equality

Cambridge, Belknap Press, 288 pages

ISBN 978-0674273559

Junaid Jahangir [MacEwan University, Canada]

The issue of inequality has received attention from top economists, whose voices have been projected through the book edited by Blanchard and Rodrik (2021). The editors make a strong case by arguing that economists should be at the forefront of combating inequality instead of making the usual naysaying arguments that "we can't afford it", "we don't have enough evidence" or that "incentives will be distorted" (p. xx). In a Canadian context, Osberg (2018) emphasizes addressing inequality as a pressing issue, stating that the Top 1% instigate narratives in their favour through think tanks and policy institutes. He proposes policies to combat inequality including guaranteed annual income, raising the top tax rates, and supporting higher wages. However, Piketty seems to have had the greatest impact on highlighting the issue of inequality, especially in mainstream newspapers, where his work has been challenged through arguments including that "inequality did not rise" and that "inequality does not matter" (Grisold and Theine, 2020). Piketty (2021) counters the mainstream, arguing that if we keep stating that it is impossible to make the richest individuals pay, we run the risk of future rebellions. In making this stand, he expresses that "there is no universal law of economics" and that everyone must "draw their own conclusions without allowing themselves to be intimidated by the well-argued opinions of others" (p. 42).

Piketty (2022) sustains this line of thought by providing an abridged reflection of his voluminous work over the last two decades. In line with Earle, Moran, and Ward-Perkins (2017), he states that economic issues are too important to be relinquished to a "small class" of experts (p. viii). In the introduction, he highlights that capitalism is connected to the "exploitation of natural resources" and "colonial domination" (p. 3). However, he adds that while it is easy to critique existing institutions that stoke inequality and oppression, it is much harder to agree on "alternative institutions" (p. 11). Additionally, in terms of combating inequality, he highlights access to education and healthcare, progressive taxes and wealth redistribution, and power sharing with employees in "business enterprises" (p. 12). Overall, he is against ceding economic issues to experts, links capitalism with exploitation, and supports access to education and healthcare, progressive taxes and wealth redistribution, and power sharing with firm employees to combat inequality.

In Chapter 1, he critiques the concept of the GDP, stating that it does not account for the depletion of natural resources or the inequality of income distribution and that even measures that account for them are problematic, as they provide the impression that damage can be "counterbalanced" by money (p. 23-24). He adds that a majority of those who contribute the greatest to carbon emissions reside in North America, whereas a majority of those who contribute the least reside in Sub-Saharan Africa and South Asia where they experience the

brunt of global warming (p. 25). By highlighting this differential impact of global warming, he argues that addressing climate change necessitates combating inequality (p. 26). Finally, he advises against the use of like the indices Gini coefficient and instead favours measures like the income proportion that goes to the bottom 50% or the top 1% (p. 29). Overall, he critiques measures that give the impression that damage can be balanced by money, indices like the Gini coefficient, the inequality in the impact of global warming, and argues that addressing climate change necessitates combating inequality.

In Chapter 2, he states that while the wealth of the middle class is concentrated in housing, the rich also have financial assets and that access to such resources allows the latter to influence politics, media, and think tanks to achieve favourable policies (p. 39-40). In Chapter 3, he states that the prescriptions of Adam Smith to uphold property rights, keep low taxes, and have balanced budgets were followed more by China than the U.K. in the 18th century (p. 52). This is because the latter accumulated significant public debt for being in a perpetual state of war and imposed taxes that accounted for 6% to 8% of the national income, which facilitated military innovations and colonial domination (p. 52-55). On free trade, he states that it was only after developing comparative advantage in the textile industry by banning imports of Indian textiles in the 18th century that the U.K. pushed towards free trade and that the same protectionist strategy was followed by Japan in the 19th century, South Korea and Taiwan in the mid 20th century, and China in the late 20th century (p. 58, 59). He adds that free trade discourse allows advanced economies to make developing countries dependent on them, as in the case of China that has made several Asian and African countries dependent on it (p. 59).

Generally, he reiterates that the wealth in western economies has been built on slavery, colonialism, and extraction of natural resources (p. 48-49). Continuing with this line of thought in Chapter 4, he highlights that reparations for slavery were paid to slaveholders and not slaves and that this payment was financed by public debt in the U.K. (p. 68, 76). Similarly, in the next chapter, he states that the French Revolution, while eliminating the privileges of the nobility, retained their property rights, as they were paid about 15% of the national income financed by taxes and public debt (p. 97, 99). Moreover, he states that reparations alone are not enough and that the damage of colonialism should be comprehensively addressed by reducing inequality through ensuring equitable access to education, employment, and property (p. 94). In this regard, he proposes the idea that every citizen on the planet should have access to tax revenues imposed on the world's billionaires, as the prosperity of the rich would not exist without the exploitation of the world's natural and human resources (p. 94). Overall, he emphasizes the impact of the wealthy on extracting favourable policies, questions the prescriptions of low taxes, balanced budgets, and free trade that the colonial powers avoided, reiterates that wealth is based on the exploitation of the world's natural and human resources, and proposes equitable access to education, employment, and property for the world's citizens funded by taxes on the world's billionaires.

In Chapter 5, he emphasizes the co-management system where half the seats in the board of directors belong to employee representatives, as they are invested in the firm's long-term viability in comparison to shareholders (p. 113-114). In Chapter 6, he states that inequality was reduced in the Western world between 1914 and 1980 through progressive taxes on income and inheritance, and the welfare state that invested in education, healthcare, pensions, public transportation, and employment insurance (p. 121-122), He highlights that confiscatory tax rates of 80% to 90% not only compel the wealthy to curb their conspicuous consumption but also create conditions against generous remunerations of top executives (p. 137-139). He adds

that there is not much of a relationship between high executive compensation and their economic performance and that their exorbitant compensation has a negative impact on the salaries of the workers below them (p. 139). Furthermore, he justifies progressive taxes by arguing that they do not discourage "innovation or productivity" and reiterates that it is education and equality that lead to development and not private property and inequality (p. 139). Overall, he highlights addressing income and wealth inequality through public investment in education and healthcare, the co-management system in firms, and progressive confiscatory taxes, which he justifies by arguing that such rates curb conspicuous consumption, that they do not impede productivity and innovation, that exorbitant executive compensation is not connected to economic performance, and that such compensation has a negative impact on the salaries of workers.

In Chapter 7, he critiques state socialism, stating that both "state ownership of the means of production and centralized planning" have failed (p. 155). Instead, he emphasizes the alternative of decentralized, democratic, and participatory socialism based on confiscatory tax rates, having employees on board of directors, and the provision of education and healthcare "beyond commercial logic" (p. 156-157, 165). Additionally, he proposes the idea of minimal inheritance for all funded by progressive wealth and inheritance taxes in conjunction with "basic income and guaranteed employment", which would empower recipients to "reject certain jobs, buy an apartment, or create a small business" (p. 160, 162, 164). Overall, he critiques state socialism and instead supports democratic, decentralized, and participatory socialism that is based on confiscatory tax rates, having employees on board of directors, the provision of education and healthcare, basic income, guaranteed employment, and inheritance for all funded by progressive wealth and inheritance taxes.

In Chapter 8, he states that encouraging women to pursue the same lifestyle as men, who earn high salaries but have limited time for family and civic participation and who contribute to consumerism and environmental damage, is not a solution to the issue of discrimination (p. 188-189). In Chapter 9, he projects the Washington Consensus that rests on budget austerity, liberalization, deregulation, and free flow of capital as neocolonialism (p. 207-208). He adds that tariffs can be justifiably reduced if they are replaced by wealth taxes, higher taxes on top incomes, and corporate taxes on multinational companies (p. 208). Additionally, he reiterates that world citizens should have access to healthcare and education funded by taxes on multinational companies and billionaires, as the wealth of the rich would not exist without the labour and resources of the poor countries (p. 215, 216). Overall, he critiques addressing discrimination through strategies that instigate consumerism and environmental damage, rejects budget austerity, liberalization, deregulation, and free flow of capital, and instead supports universal access to healthcare and education funded by taxes on multinational companies and billionaires.

In Chapter 10, he critiques Chinese socialism as "statist and authoritarian" that is at odds with his proposition of democratic, decentralized, ecological, and participatory socialism (p. 226). Additionally, he states that the only limit to monetary policy is inflation and that it can be used to achieve full employment, reduce inequality, and save the planet, but that it has often fuelled higher prices in the stock and the real estate markets, thereby adding to the wealth of the rich (p. 240-241). Overall, he critiques Chinese socialism, highlights the limits of monetary policy, and upholds his vision of participatory socialism that rests on progressive taxes, power sharing in firms, guaranteed employment, access to healthcare and education, and inheritance for all (p. 237).

To recapitulate, Piketty argues against ceding economic issues to experts and argues that addressing climate change necessitates combating inequality. He links capitalism with exploitation, reiterates that wealth is based on the exploitation of the world's natural and human resources, and questions the neoliberal prescriptions of low taxes, balanced budgets, free trade, budget austerity, liberalization, deregulation, and free flow of capital. He critiques both state socialism and Chinese socialism, and instead supports democratic, decentralized, and participatory socialism that is based on confiscatory tax rates, having employees on board of directors, the provision of universal education and healthcare, basic income, guaranteed employment, and inheritance for all to address income and wealth inequality. Overall, he favours progressive wealth and inheritance taxes, and taxes on multinational companies and billionaires to allow all the citizens of the world access to education and healthcare, based on the argument that wealth is collective and that the riches of the wealthy are dependent on the labour of the poor and the resources of the poor countries. Thus, Piketty offers a narrative that rejects both capitalism and state socialism and instead upholds a democratic, decentralized, ecological, and participatory socialism.

References

Blanchard, O., and Rodrik, D. (2021) *Combating Inequality*. Cambridge: The MIT Press.

Earle, J., Moran, C., and Ward-Perkins, Z. (2017) *The Econocracy*, Manchester: Manchester University Press.

Grisold, A., and Theine, H. (2020) "Now what exactly is the problem?" Media coverage of economics inequalities and redistribution policies: The Piketty Case, in: *Journal of Economic Issues*, 54(4): 1071-1094.

Osberg, L. (2018) *The Age of Increasing Inequality,* Toronto: Lorimer.

Piketty, T. (2021) *Time for Socialism*, New Haven and London: Yale University Press.

Author contact: junaid@ualberta.ca

SUGGESTED CITATION:
Junaid Jahangir, "Review of Thomas Picketty's '*A Brief History of Equality*'", *real-world economics review*, issue no. 103, 31 March 2023, pp. 128-131, http://www.paecon.net/PAEReview/issue103/Jahangir_Review

You may post and read comments on this paper at http://rwer.wordpress.com/comments-on-rwer-issue-no-103/

Book Review

John Komlos

***Foundations of Real-World Economics,* 3rd Edition**

Cambridge, Routledge, 420 pages
ISBN 978-1032004846

Alan Freeman

John Komlos has updated his justly-reputed book with three additional chapters which I will discuss at the end of this review. The new edition is a valuable extension that reenforces the original with reference to modern events, justifying the cost of a second copy if you already have the first; if you haven't, it's a bargain.

Komlos' persistent and well-evidenced criticism of mainstream goes to the heart of its primary failure: namely it does not describe reality. What it predicts simply doesn't occur, as, it's fair to say, is well-known. Yet in a certain sense, it's not the problem. Scientists make predictive failures all the time, and indeed, this is how science progresses: it constantly tests its theories to see if their predictions match what is observed, and when the theory doesn't match reality, they change the theory.

So while the many empirical mismatches between economic theory and reality do tell us that it's pretty foolish to rely on the theory, these failures don't actually count as evidence that it is incapable of getting anything right. In fact, perversely, this makes economics attractive as a career option: since there is so much wrong with it, a lot of money is to be made by peddling fixes. Indeed, you can even win "Nobel Prizes" for this kind of activity, which usually amounts to developing sticking plaster for gaping wounds. It is noteworthy that the demand for economists is countercyclical, which is to say, the worse the results of employing them, the greater is the demand to know why.

A rather more important point is this: economics rarely if ever changes its mind when it encounters such failures. Now, there are scientists like that, as Thomas Kuhn first pointed out. But in pretty well all true sciences, there always comes a point, no matter how strong the resistance, where a badly-performing theory is discarded and replaced by a better one. This is not so for economics.

To the contrary, as the aftermath of 2008 shows, it doubles down on the mistakes that produce the bad results, such as insisting that finance is a productive activity. This is a danger sign that something more fundamental is wrong with the way economic theory is managed, which has not only divorced it from the 'Real World' but immunised it against ever reconnecting. Komlos's revised final chapter is distinctively more pessimistic about the prospects of fixing this, as am I. But which are the mistakes that produce the bad results?

This is what Komlos—who comes from a scientific background—sets out to identify. The link to 'reality' is in the nature of these mistakes, which are not merely 'wrong' in the sense of, say,

erroneously claiming that the Icelandic economy is in a splendid state just before it collapsed, or predicting that Shock Therapy would put the Russian economy on a sound footing. The problem is the *assumptions that lead* to these mistakes. The point is that the assumptions, themselves, are *unreal.* They are built into the fabric of economics, and so cannot but lead it astray. They include ideas like the fallacy that consumers make rational choices on what to buy based on an internalised set of psychological preferences, or the Neoplatonic notion of 'perfect competition', which is incompatible with any observed reality. John Weekes once summed up this approach with the observation that for economics, a horse is an imperfect Unicorn. Discarding these assumptions would demand of economics such a comprehensive overhaul that it is, in its current state, incapable of so doing.

Komlos scores on this front too: he recommends remedies. He endorses what he terms 'humanistic economics' which "implies the vision that a kinder and more just economic system is possible, one that is embedded in a truly democratic society that not only empowers people but enables them to live their daily lives with less uncertainty, less manipulation, less taking advantage of people's weaknesses, and less fear that their lives could collapse like a house of cards."

The most intriguing part of the new book comes at the end, in the three new chapters which make the whole book worth rereading. The titles say it all." Economists' Mistakes Lead to Right-Wing Populism Plus an Insurrection"; "Hidden Racist Elements in Blackboard Economics" and "The Covid-19 Pandemic Exposed the Need for a Black-Swan-Robust Economy". The mistakes that Economists inflict go beyond the arid world of academic theory; they not only fail to understand the Real World, but happily participate in destroying it. Every undergraduate should be made to know, before even embarking on this perilous and pernicious career, both the havoc they could wreak unless they stand up to their orthodox masters, and the good they might do by resisting.

Author contact: Alan.Freeman@umanitoba.ca

SUGGESTED CITATION:
Alan Freeman, "Review of John Komlos's '*Foundations of Real-World Economics*', 3rd Edition", *real-world economics review*, issue no. 103, 31 March 2023, pp. 132-133,
http://www.paecon.net/PAEReview/issue103/Freeman

End Matter

ISSN 1755-9472

EDITOR: Edward Fullbrook,
ASSOCIATE EDITOR: Jamie Morgan
PAST CONTRIBUTORS: James Galbraith, Frank Ackerman, André Orléan, Hugh Stretton, Jacques Sapir, Edward Fullbrook, Gilles Raveaud, Deirdre McCloskey, Tony Lawson, Geoff Harcourt, Joseph Halevi, Sheila C. Dow, Kurt Jacobsen, The Cambridge 27, Paul Ormerod, Steve Keen, Grazia Ietto-Gillies, Emmanuelle Benicourt, Le Movement Autisme-Economie, Geoffrey Hodgson, Ben Fine, Michael A. Bernstein, Julie A. Nelson, Jeff Gates, Anne Mayhew, Bruce Edmonds, Jason Potts, John Nightingale, Alan Shipman, Peter E. Earl, Marc Lavoie, Jean Gadrey, Peter Söderbaum, Bernard Guerrien, Susan Feiner, Warren J. Samuels, Katalin Martinás, George M. Frankfurter, Elton G. McGoun, Yanis Varoufakis, Alex Millmow, Bruce J. Caldwell, Poul Thøis Madsen, Helge Peukert, Dietmar Lindenberger, Reiner Kümmel, Jane King, Peter Dorman, K.M.P. Williams, Frank Rotering, Ha-Joon Chang, Claude Mouchot, Robert E. Lane, James G. Devine, Richard Wolff, Jamie Morgan, Robert Heilbroner, William Milberg, Stephen T. Ziliak, Steve Fleetwood, Tony Aspromourgos, Yves Gingras, Ingrid Robeyns, Robert Scott Gassler, Grischa Periono, Esther-Mirjam Sent, Ana Maria Bianchi, Steve Cohn, Peter Wynarczyk, Daniel Gay, Asatar Bair, Nathaniel Chamberland, James Bondio, Jared Ferrie, Goutam U. Jois, Charles K. Wilber, Robert Costanza, Saski Sivramkrishna, Jorge Buzaglo, Jim Stanford, Matthew McCartney, Herman E. Daly, Kyle Siler, Kepa M. Ormazabal, Antonio Garrido, Robert Locke, J. E. King, Paul Davidson, Juan Pablo Pardo-Guerra, Kevin Quinn, Trond Andresen, Shaun Hargreaves Heap, Lewis L. Smith, Gautam Mukerjee, Ian Fletcher, Rajni Bakshi, M. Ben-Yami, Deborah Campbell, Irene van Staveren, Neva Goodwin, Thomas Weisskopf, Mehrdad Vahabi, Erik S. Reinert, Jeroen Van Bouwel, Bruce R. McFarling, Pia Malaney, Andrew Spielman, Jeffery Sachs, Julian Edney, Frederic S. Lee, Paul Downward, Andrew Mearman, Dean Baker, Tom Green, David Ellerman, Wolfgang Drechsler, Clay Shirky, Bjørn-Ivar Davidsen, Robert F. Garnett, Jr., François Eymard-Duvernay, Olivier Favereau, Robert Salais, Laurent Thévenot, Mohamed Aslam Haneef, Kurt Rothschild, Jomo K. S., Gustavo Marqués, David F. Ruccio, John Barry, William Kaye-Blake; Michael Ash, Donald Gillies, Kevin P.Gallagher, Lyuba Zarsky, Michel tmoerBauwens, Bruce Cumings, Concetta Balestra, Frank Fagan, Christian Arnsperger, Stanley Alcorn, Ben Solarz, Sanford Jacoby, Kari Polanyi, P. Sainath, Margaret Legum, Juan Carlos Moreno-Brid, Igor Pauno, Ron Morrison, John Schmitt, Ben Zipperer, John B. Davis, Alan Freeman, Andrew Kliman, Philip Ball, Alan Goodacre, Robert McMaster, David A. Bainbridge, Richard Parker, Tim Costello, Brendan Smith, Jeremy Brecher, Peter T. Manicas, Arjo Klamer, Donald MacKenzie, Max Wright, Joseph E. Stiglitz. George Irvin, Frédéric Lordon, James Angresano, Robert Pollin, Heidi Garrett-Peltier, Dani Rodrik, Marcellus Andrews, Riccardo Baldissone, Ted Trainer, Kenneth J. Arrow, Brian Snowdon, Helen Johns, Fanny Coulomb, J. Paul Dunne, Jayati Ghosh, L. A Duhs, Paul Shaffer, Donald W Braben, Roland Fox, Stevan Harnad, Marco Gillies, Joshua C. Hall, Robert A. Lawson, Will Luther, JP Bouchaud, Claude Hillinger, George Soros, David George, Alan Wolfe, Thomas I. Palley, Sean Mallin, Clive Dilnot, Dan Turton, Korkut Ertürk, Gökcer Özgür, Geoff Tily, Jonathan M. Harris, Jan Kregel, Peter Gowan, David Colander, Hans Foellmer, Armin Haas, Alan Kirman, Katarina Juselius, Brigitte Sloth, Thomas Lux, Luigi Sapaventa, Gunnar Tómasson, Anatole Kaletsky, Robert R Locke, Bill Lucarelli, L. Randall Wray, Mark Weisbrot, Walden Bello, Marvin Brown, Deniz Kellecioglu, Esteban Pérez Caldentey, Matías Vernengo, Thodoris Koutsobinas, David A. Westbrook, Peter Radford, Paul A. David, Richard

Smith, Russell Standish, Yeva Nersisyan, Elizabeth Stanton, Jonathan Kirshner, Thomas Wells, Bruce Elmslie, Steve Marglin, Adam Kessler, John Duffield, Mary Mellor, Merijn Knibbe, Michael Hudson, Lars Pålsson Syll, Jane D'Arista, Ali Kadri, Egmont Kakarot-Handtke, Ozgur Gun, George DeMartino, Robert H. Wade, Silla Sigurgeirsdottir, Victor A. Beker, Pavlina R. Tcherneva, Dietmar Peetz, Heribert Genreith, Mazhar Siraj, Hazel Henderson, Nicolas Bouleau, Geoff Davies, D.A. Hollanders, Richard C. Koo, Jorge Rojas, Marshall Auerback, Fernando García-Quero, Fernando López Castellano, Robin Pope and Reinhard Selten, Patrick Spread , Wilson Sy, Fred Moseley, Shimshon Bichler, Johnathan Nitzan, Nikolaos Karagiannis, Alexander G. Kondeas, Roy H. Grieve, Samuel Alexander, Asad Zaman, L. Frederick Zaman, Avner Offer, Jack Reardon, Yinan Tang, Wolfram Elsner, Torsten Heinrich, Ping Chen, Stuart Birks, Dimitrios Koumparoulis, Arne Heise, Mark Jablonowski, Carlos Guerrero de Lizardi, Norbert Häring, William White, Jonathan Barzilai, David Rosnick, Alan Taylor Harvey, David Hemenway , Ann Pettifor, Dirk Helbing, Douglas Grote, Brett Fiebiger, Thomas Colignatus, M. Shahid Alam, Bryant Chen, Judea Pearl, John Pullen, Tom Mayer, Thomas Oechsle, Emmanuel Saez, Joseph Noko, Joseph Huber, Hubert Buch-Hansen, Brendan Sheehan, C P Chandrasekhar, Heikki Patomäki, Romar Correa, Piet-Hein van Eeghen, Max Koch, John Robinson, Oscar Ugarteche, Taddese Mezgebo, Donald Katzner, Crelis F. Rammelt, Phillip Crisp, John B. Benedetto, Alicia Puyana Mutis, Leon Podkaminer, Michael Kowalik, Mohammad Muaz Jalil, José A. Tapia, Robert W. Parenteau, Alan Harvey, C. T. Kurien, Ib Ravn, Tijo Salverda. Holger Apel, John Jeffrey Zink, Severin Reissl, Christian Flamant, Rainer Kattel, Amit Bhaduri, Kaustav Banerjee, Zahra Karimi Moughari, David R Richardson, Emil Urhammer, Michel Gueldry, Rüya Gökhan Koçer, Hee-Young Shin, Kevin Albertson, John Simister, Tony Syme, Geoff Tily, Ali Abdalla Ali, Alejandro Nadal, Steven Klees, Gary Flomenhoft, Bernard C. Beaudreau, William R. Neil, Ricardo Restrepo Echavarría, Carlos Vazquez, Karen Garzón Sherdek, Paul Spicker, Mouvement des étudiants pour la réforme de l'enseignement en économie, Suzanne Helburn, Martin Zerner, Tanweer Akram, Nelly P. Stromquist, Sashi Sivramkrishna, Ewa Anna Witkowska, Ken Zimmerman, Mariano Torras, C.P. Chandrasekhar, Thanos Skouras, Diego Weisman, Philip George, Stephanie Kelton, Luke Petach, Jørgen Nørgård, Jin Xue, Tim Di Muzio, Leonie Noble, Kazimierz Poznanski, Muhammad Iqbal Anjum, Pasquale Michael Sgro, June Sekera, Michael Joffe, Basil Al-Nakeeb, John F. Tomer, Adam Fforde, Paulo Gala, Jhean Camargo, Guilherme Magacho, Frank M. Salter, Michel S. Zouboulakis, Prabhath Jayasinghe, Robert A. Blecker, Isabel Salat, Nasos Koratzanis, Christos Pierros, Steven Pressman, Eli Cook, John Komlos, J.-C. Spender, Yiannis Kokkinakis, Katharine N. Farrell, John M. Balder, Blair Fix, Constantine E. Passaris, Michael Ellman, Nuno Ornelas Martins, Jason Hickel, Eric Kemp-Benedict, Sivan Kartha, Peter McManners, Richard B. Norgaard, William E. Rees, Joachim H. Spangenberg, Lia Polotzek, Per Espen Stoknes, Imad A. Moosa, Salim Rashid, Richard Vahrenkamp, Dimitri B. Papadimitriou, Michalis Nikiforos, and Gennaro Zezza, Abderrazak Belabes, Phil Armstrong, Bruno Bonizzi, Annina Kaltenbrunner, Jo Michell, Dirk H. Ehnts, Maurice Höfgen, Richard Murphy, Louis-Philippe Rochon, Malcolm Sawyer, Jan Toporowski, Kenneth Austin, Duncan Austin, Kalim Siddiqui, Arthur M. Diamond, Jr., Theodore P. Lianos, John E. Coulter, Gerald Holtham, Jaehee Choi, Girol Karacaoglu, Geoff Crocker, Andri W. Stahel, Gregory A. Daneke, Joaquim Vergés-Jaime, Hardy Hanappi. Rosemary Batt, Patrick Pobuda, Ulrich Thielemann, Tanja von Egan-Krieger, Steve Roth, Ron Wallace.

Articles, comments on and proposals for should be sent to the editor at
pae_news@btinternet.com

Subscriptions to this email journal are free to all members of the World Economics Association (WEA). WEA memberships are free at http://worldeconomicsassociation.org/

Back issues of this journal and material related to the PAE movement are available at www.paecon.net.

To subscribe to this journal go **here** (http://www.feedblitz.com/f/f.fbz?Sub=332386)

To unsubscribe to this journal go to

http://www.feedblitz.com/f/f.fbz?EmailRemove=_NDUzNDkzMHwzMzIzODZ8cGFlX25ld3NAYnRpbnRlcm5ldC5jb218MTA5OTU4_