real-world economics review

Please click here to support this journal and the WEA

- Subscribers: 26,369 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 World Social and Economic Review

back issues

Issue no. 83
20 March 2018

Ten years after the crisis: a lost decade?
Steven Pressman and Robert Scott

The great marginalization: why twentieth century economists neglected inequality
Eli Cook

Game Theory
On the current state of game theory
Bernard Guerrien

Why game theory never will be anything but a footnote in the history of social science
Lars Pålsson Syll

Employment
The creation of jobs
Michael Joffe

Employment in a just economy
John Komlos

Business Studies
Managing the engines of value-creation
J.-C. Spender

The effect of academic business studies in Germany and America in the modern era
Robert R. Locke

Does the maximization principle break down during recessions?
Philip George

From Bologna to Athens: the political economy of reforms in Europe
Yiannis Kokkinakis

Book review essay
The Vienna school of ecological economics
Katharine N. Farrell

Board of Editors, past contributors, submissions, etc.

support this journal and the WEA

visit the RWER Blog
Ten years after the crisis: a lost decade?
Steven Pressman and Robert Scott [Colorado State University and Monmouth University, USA]

Abstract
The Great Recession was the most significant economic downturn since the Great Depression. This paper discusses the primary causes of the Great Recession and studies whether these problems were resolved or are lying dormant at the periphery of the system waiting to wreak havoc again. We focus on the increasing financialization of the economy and the current state of household finances. Has the ten years following the Great Recession been a lost decade for American households?

JEL codes D14, E21, E44, G18, G28

Key words deregulation, financialization, housing bubble, household debt, income inequality, Glass-Steagall Act, Great Recession

"The more things change, the more they stay the same" Bon Jovi (as well as Jean-Baptiste Alphonse Karr).

1. Introduction

The Great Recession was the most significant economic downturn since the Great Depression. Understanding its primary causes and consequences is of immense importance. This paper discusses the causes of the Great recession, but more importantly, it studies whether the problems leading to the Great Recession have been resolved, or whether they are still lurking below the surface of the United States economy.

There is currently no agreement as to when the Great Recession began. Some take December 2007, the official NBER dating of the recession, as the official starting date. On this approach, the recession lasted until June 2009, making it the longest recession since the end of World War II. It also had the most severe financial consequences of all post-War recessions. According to the Federal Reserve (Rich, 2013),

“Home prices fell approximately 30 percent, on average, from their mid-2006 peak to mid-2009, while the S&P 500 index fell 57 percent from its October 2007 peak to its trough in March 2009. The net worth of US households and non-profit organizations fell from a peak of approximately $69 trillion in 2007 to a trough of $55 trillion in 2009.”

Others date the start of the Great Recession with the fall of Lehman Brothers on September 15, 2008. A good case can be made that this precipitated the subsequent stock market plunge, massive layoffs, sharply falling home prices, and government bailouts that we associate today with the Great Recession. Once Lehman failed, AIG could not make good on the loans made by other institutions to Lehman that it had insured. This put AIG at risk of collapse, as well as other financial institutions that had lent money to Lehman. As everyone sought safety, this led to large withdrawals from financial institutions, which were then forced to curtail their lending. A deep recession became inevitable.
We prefer another starting date for the Great Recession, February 2008, when the Auction Rate Securities (ARS) market collapsed, with more than 75% of auctions failing. Estimates put the value of this market at $330 billion at the time; so around $250 billion in assets were frozen in the US, rendering many investors illiquid (Lee, 2008). Seven years later, $50 billion of ARSs were still frozen (Doherty, 2015).

ARSs were developed in the early 1980s by Ronald Gallatin of Lehman Brothers. The first ARS was registered with the SEC in July 1984, and the first auctions took place a few months later. As the name implies, they carry a variable interest rate that is set periodically by auction. These securities typically receive AAA ratings before being auctioned off because security raters look at default probability when evaluating securities but not the possibility of auction or market failure. ARSs were “sold” to customers of financial institutions as an alternative to money-market funds – they provided liquidity and a slightly higher rate of interest. With a minimum investment of $25,000, ARSs were held by many upper middle-class households. They were intended to finance retirement, a child’s college education, or to buy a home.

These securities can probably be best described, following Keynes (1936, Ch. 12), as a game of Old Maid. There is a regular auction to see who gets the Old Maid card. Those winning the auction get slightly higher rate of return on their money. However, the game can end at any point. In this case, whoever holds the Old Maid is stuck because they cannot get their cash. They must wait until the game starts up again, or until a market develops again for these securities. If no market develops, they must wait until the underlying securities get paid off, which can take many years.

Whatever date is chosen as the start of the Great Recession, it is clear that something happened in early and mid-2008 causing severe damage to the US economy – unemployment at 10%, home values falling by more than 25% nationwide, and historic bank bailouts. This paper looks at the Great Recession around ten years after it began. It first examines primary causes of this event; then it argues that the problems leading to the Great Recession have not been solved. The final section concludes with some policy solutions.

2. What led to the Great Recession?

Just as there is little agreement on a starting date, there is little agreement on the causes of the Great Recession. A large number of studies (Blinder, 2013; Jarrow, 2011; Mian and Sufi, 2014; National Commission on the Causes of the Financial and Economic Crisis in the United States, 2011; Roubini and Mihn, 2010; Stiglitz, 2009; van Treeck and Sturn, 2012, to name just a few) have sought to determine what led to the economic and financial crisis that started ten years ago. Many suspects have been identified – lack of adequate government regulation, a worldwide saving glut, mortgage fraud, sub-prime and predatory lending, failures of corporate governance and risk management, a breakdown in accountability and ethics, and rising inequality. So many guilty parties have been identified that it feels like Agatha Christie’s Murder on the Orient Express – spoiler alert, everyone did it.
In what follows we winnow down the main causes to a few related culprits – *stagnating household incomes* over a long period of time along with *rising household debt; housing problems; and government regulatory failure* regarding large financial institutions and shadow banks.

Let’s start by looking at households and household income. Figure 1 below shows real median household income and per capita real Gross Domestic Product (GDP) since 1984. A few things stand out here. Median incomes decline during recessions (1991, 2001 and 2008) and then recover during the subsequent expansion. Over the long term, median incomes rose through the late 1980s and most of the 1990s, but stagnated thereafter. US median household income in 2016 was $59,039; in real terms this is nearly the same as 1999. To put this into a broader historical context, real median incomes grew during the four decades following World War II, and people came to expect that this would continue when making financial decisions. Starting in the late 1980s, we have seen only cyclical variations. The assumption of constantly rising living standards no longer holds true.

Equally concerning is the fact that per capita GDP has increased over 70% since 1984, while median incomes rose less than 20%. The slowdown of median income growth is *not* due to a lack of productivity, or reduced labor supply, but rather indicates a distributional problem. Another way to see this is through the difference between the growth in worker productivity and the growth in worker compensation. These two growth rates tracked each other closely until the 1970s, when they first began to diverge. Then they began to move apart more quickly, as well as continuously, with workers getting only a small fraction of the gains due to their improved productive capacity (see Figure 2).

**Figure 1** Real median incomes in United States and real GDP per capita, 1984-2016

![Figure 1](image_url)
Things are even worse than the standard data show. One problem is the rise of household debt, something that we have focused on in previous work (Pressman and Scott, 2009; Scott and Pressman, 2011). Some debt was incurred during recessionary times, when government support (such as unemployment insurance) was inadequate. Some was incurred to deal with rising inequality (Kumhof et al., 2015). With wages falling for many people, households borrowed to pay fixed and relatively fixed expenses (such as rent or mortgages, food, car payments, insurance, and utilities). Also, contemporary consumption-based economies encourage consumption as part of identity. As previous non-necessary goods become en vogue in society (e.g., expensive coffee drinks) people engage in what Thorstein Veblen (1899) called pecuniary emulation – often beyond their financial limits. Finally, some debt undoubtedly stemmed from bad luck – a health problem, parents who could not or would not help pay for college, or an investment decision that turned out bad only in hindsight (Frank, 2016).

Whatever the cause, debt must be repaid with interest. Looking at just the interest payments (and ignoring principal repayment), we previously estimated that government poverty rates were underestimated by 8% because they did not include lost income due to interest payments on consumer debt. We also found that, when interest payments on consumer debt get subtracted from income, income inequality was underestimated by 7.2% (Pressman and Scott, 2009). Likewise, the lost income from interest payments on consumer debt reduced the size of the middle class by 11.7%, or 2.1% of US households were “squeezed” out of the

---

**Figure 2** Hourly compensation (wages and benefits) and productivity growth (per hour worked) in the United States, 1948-2016

middle class due to income going to make interest payments rather than purchase goods and services (Scott and Pressman, 2011).

A second problem with the standard measure of household income is that it assumes households remain the same over time. According to the US Census Bureau, a household is officially defined as one or more people living together under the same roof. The make-up of these households can and do change over time due to economic circumstances and sociological factors. While this is not the place to analyze the causes of this change, we want to point out that there are consequences of this for estimates of median household income.

Consider a married couple with each making less than the median income, but whose combined income puts them above the median. If they divorce, then one income above the mid-point gets replaced with two incomes below the median. As a result, the new median becomes the income level that was just below the prior median income. If this happens hundreds of thousands of times, real median income can fall substantially.

Or consider a student who graduates from college. When jobs are plentiful and incomes reasonably good, the graduate can set up their own household and pay rent, utilities, etc. with their income (maybe a little help from their parents). Since young graduates are typically single and usually earn less money when starting their working career, they are likely to become a single-person household and earn less than the median income. When graduates form their own household when leaving college, this pushes down the median income. In contrast, when graduates cannot find a good job and live in their parents’ home, or what Brunnermeier et al. (2016, pp. 110, 244) call “Hotel Mama,” it increases income for many households because the income from the graduate living at home gets added to the income of other household members. Median incomes therefore rise.

This latter change is not just some theoretical curiosity. Fry and Passel (2014) report a sharp increase in multi-generational family households between 2000 and 2012. For the population as a whole, multi-generational households increased from 15.1% to 18.1% (bringing it back to 1950s levels) while for individuals aged 25 to 34, the increase was from 15.8% living in multi-generational families to 23.1%, returning it to 1940s levels. Moreover, they find this change beginning in earnest in 2008, at the outset of the Great Recession.

There is a third problem with the standard economic figures on household income. Seeking to maintain their standard of living, many households have increased their work effort. More family members are working and are working additional hours, additional days and additional jobs. The extra money gets counted in household income. But there is another side to greater work effort – it comes with higher costs. And these costs do not get counted in standard economic measures that are supposed to track living standards. Living costs rise due to such things as additional transportation requirements, and people eating out more because they lack the time and energy to cook meals at home. Most important of all, there are additional child care costs for many families when all the adults are working more. A family with a four-year old and an infant spend, on average, $18,000 per year on child care; this constitutes around one-third of the household income for a typical American family (Glynn and Corley, 2016).

---

2 We define “middle-class households” as those with adjusted household income between 50% and 150% of median adjusted household income.
The cost of additional work effort may go even beyond the monetary cost. Households squeezed by lower income, greater debt and additional work lack time and the mental bandwidth to make good financial decisions (Mullainathan and Shafir, 2013). The financialization of everyday life (see Bryan et al., 2012) requires people to be experts in their financial affairs within a system that is complex and uncertain (e.g., the transition from pensions to 401Ks). Financial institutions are able to take advantage of this financial knowledge gap, which is where regulatory failure also comes into play. Deregulating the financial industry enabled firms to engage in risky, deceptive and illegal behavior, setting the stage for the housing bubble. Financial institutions pushed risky investments on unsuspecting middle-class households looking to maintain their economic condition and status in the face of economic slowdowns and rising inequality. For example, they promised higher returns on an investment that they claimed was completely liquid (ARSs), and they promised borrowers that their homes could only go up in price and that mortgages could always be refinanced with the gains from higher home prices.

The Financial Crisis Inquiry Report (National Commission on the Causes of the Financial and Economic Crisis in the United States, 2011) singles out the Federal Reserve, and Alan Greenspan in particular, for failing to do their job. Created to protect the public and ensure the safety and soundness of the US financial system, the Fed refused to set reasonable lending standards. It also failed to heed warnings about predatory lending – including warnings from Fed Governor Ed Gramlich (2007), while he was on his deathbed. Lack of regulation led to liar loans and NINJA (No Income, No Job, No Assets, no problem) loans, and to the exotic mortgages (some with interest-only payments and some where the principal grew over time rather than fell) with low teaser rates that homeowners could only pay if the value of their home increased sharply in value so that the mortgage could be refinanced when higher rates were about to kick in. These mortgages were then packaged, given AAA ratings (by agencies needing the business and ignoring systemic risk), and then sold or rented (as ARSs) to unsuspecting (or ill-informed) investors (see Pozsar, 2008).

3. Where are we now?

Casting blame is always fun; but that is not our purpose here. We wish to address a more serious issue – whether the problems leading to the Great Recession have been solved, or whether they lurk below the surface of the US economy. We fear the serious problems remaining, and we are concerned that once the US economy stops growing the next recession will also be deemed “Great,” similar to how the economic collapse of 1887-9 was called the “Great Depression” before the 1930s.

Considerable economic history backs the view that we should be worried about another economic and financial crisis. Until World War II, the US economy experienced frequent economic problems that looked just like a Great Recession or Great Depression. The panic of 1907 (Bruner and Carr, 2007) was the twelfth bank panic since 1814, and one of the worst (Calomiris and Gorton, 2000) in US history. It was so bad that private bankers (such as J.P. Morgan) could no longer solve the problems associated with a financial crisis, leading to the formation of the Federal Reserve in 1913. Nonetheless, panics arose 1914, in 1921 and again in 1929, when the great stock market crash led to the Great Depression. Over the 100-year period from the early 1800s to the early 1900s, the US economy encountered severe economic problems every 8-9 years on average.
For these reasons, we are interested in what happened 10 years ago and whether problems remain that threaten our economic future. As in the previous section, we divide these problems into three groups – stagnating household incomes (in conjunction with rising debt levels), housing, and inadequate regulation of financial institutions. This last problem is especially noteworthy because the crises in the 19th and early 20th century were financial panics that then damaged Main Street and average citizens. In contrast, the period from the end of World War II to the end of the 20th century was characterized by significant government regulation of the financial industry with few financial panics. Over time, the dominant paradigm changed from market regulation to a belief that free markets were more efficient than government regulation.

Deregulation began in the 1980s, and continued through the 1990s. The Garn-St Germain Depository Institutions Act of 1982 deregulated credit unions as well as savings and loans, and allowed non-bank banks (mortgage companies, payday lenders, and hedge funds that take in money from wealthy individuals and use this money to make loans) to exist. These “shadow banks” grew rapidly, unhampered by the many restrictions placed on banks. Their depositors or investors received higher returns on their money because they made risky loans at high interest rates. Deregulating these institutions increased the pressure to deregulate large commercial banks, so that they could compete with lightly regulated shadow banks.

The 1994 Riegel-Neal Interstate Bank Efficiency Act repealed restrictions on interstate banking. This fueled the rise of mega financial institutions that became too big to fail. Knowing that the government would have to bail them out if they were in jeopardy of going under, large banks could take on even greater risks since the downside of aggressive lending (bankruptcy) was mitigated by the implicit promise of a government bailout.

Finally, in November 1999 President Clinton signed the Gramm-Leach-Bliley bill, thereby repealing the Glass-Steagall Act of 1933. A New Deal reform, Glass-Steagall established deposit insurance and limited the risks that commercial banks could take with insured deposits. Under Glass-Steagall a bank could accept deposits and also make loans, or it could sell securities. However, it could not do both. If a bank made a loan, it had to keep that loan on its books, since it was prohibited from selling it. In addition, Glass-Steagall required that banks offer only standard fixed-rate mortgages. Its repeal led to the creation of exotic mortgages, which financial institutions then packaged together and sold off.

It seems noteworthy that a mere 9 years following repeal of Glass-Steagall we had our next major financial crisis.

3.a) Households

The good news for households is that jobs have returned. The US unemployment rate has fallen from 10% in October 2009, its peak during the Great Recession, to 4.1% in October 2017. Another bit of good news is that median household income (adjusted for inflation) has increased since 2012 by $5,708 – from $53,331 in 2012 to $59,039 in 2016 – putting it slightly above its level right before the Great Recession. The bad news is that median household income remains nearly unchanged since 1999 – resulting in no gains in living standards for typical household over a period of nearly two decades.

However, things are actually worse than this. As noted earlier in this paper, median income figures, released annually by the Census Bureau, do not take into account family size. As a
result, they ignore the large demographic changes that took place over the past decade – fewer new households being created, adult children more likely to live with their parents, even returning home after college rather than finding a good job and establishing a household of their own. Adjusting household income by family size is a standard procedure for measuring inequality and computing relative poverty rates in cross-national studies. It is also standard in measures of absolute poverty rates, such as the Orshansky (1965; 1969) definition of poverty employed in the US.

Using the Federal Reserve’s Survey of Consumer Finances (SCF) data (Federal Reserve Board of Governors 2017b), we adjusted household income using the OECD equivalence scale (see Atkinson, Rainwater and Smeeding, 1995), where each additional adult in the household counts as needing 0.5 of the income of the household to keep living standards constant and each child needs 0.7 of the income of the household head. We found that the real median adjusted household income was $33,315 in 2007, right before the start of the Great Recession; but in 2016 it was $32,220, or 3.3% lower. This decline does not show up in standard income measures. However, it indicates clearly that US households have lost ground; they are not yet back to where they were before the Great Recession.

Furthermore, these figures understate the two major financial problems plaguing many US households – debt and inadequate savings. Many people live paycheck to paycheck; nearly half (46%) of all adults say they could not come up with $400 in case of emergencies (Board of Governors of the Federal Reserve, 2016). The Federal Reserve Bank of New York (2017) reported that, in the first quarter of 2017, household debt exceeded debt levels from before the Great Recession. While these figures do not control for inflation during the past decade, it remains disturbing that debt levels are near their all-time peak. Moreover, the issue is not just debt; it is the ability to repay that debt. People are already struggling with high debt payments that they can barely afford to pay. Our debtor society is a result of loose lending practices, strict bankruptcy laws and flat real wages. What will happen when the next recession hits?

Several aspects of household debt are especially troubling. Rising college costs and the increase in college debt has become headline news. Student loan debt rose over 101% in real terms from an average of $3,789 in 2007 to $7,623 in 2016. This poses problems because this debt cannot easily be dismissed, as credit card debt can be dismissed in bankruptcy. If not paid, it will come out of Social Security payments during retirement. This debt also reduces the spending by young households as they begin their careers and adult lives, and is a main cause of the “Hotel Mama” phenomenon. Another concern is motor vehicle debt – both because auto prices have not increased substantially over the past several decades and because of auto title loans. Auto title loans use one’s motor vehicle as collateral and typically come with higher interest rates than other sources of credit. A few missed payments can result in repossession, preventing people from getting to work and putting their jobs at risk. Last, but not least, credit card debt (an expensive way to borrow) is now at record levels, exceeding $1 trillion.

Updating our past work on the impact of household debt using the latest SCF 2016 data, we find that debt-to-median income ratios rose to 0.4 in 2016, compared to 0.349 in 2007, an increase of 14.6%. Also, it should be noted that 5.39% of households filed for bankruptcy over this period, wiping out their debt and thereby reducing debt ratios.

It is clear that US households are in trouble. Real median income, adjusted for household size, has fallen 3.3% from 2007 to 2016. We can add to this decline another 3.1% due to
rising interest payments on consumer debt. With the social safety net in tatters, it is no wonder that American households are struggling and are worried about what the future will bring.

This also raises an important question – how long can households accumulate debt, paying more and more of their income as interest on past debt? Or, to use Minsky’s (1977; 1982) framework, when does speculative finance become Ponzi finance? Minsky focused on firm debt rather than household debt; his point of transition occurred when income was no longer sufficient to even pay the interest on past debt. Firms, then, had to borrow in order to pay the interest on past debt. For households we have not reached this point and probably never will – primarily because households, unlike firms, need food, clothing, and shelter in order to survive. However, we are approaching a point where households cannot purchase basic necessities and also make necessary interest payments on their past debt. This may not be the Ponzi finance that Minsky described; perhaps we need another name, something like “Lehman finance,” where households cannot sustain themselves and also pay interest on their past debt. We do need to worry about what happens when we reach this point. We also need to understand what the practical limits are to household debt levels and whether we are approaching these limits. However, these are issues beyond the scope of this paper.

3.b) Housing

History shows how important housing is when it comes to economic and financial crises. The Nordic countries experienced problems like those of the US (and elsewhere) in the 1980s and 1990s (Moe et al., 2004). Housing was also a problem in the US during the Great Depression. In many ways, the 1920s and early 2000s are similar with respect to housing. During the 1920s, US home prices rose 45%. Prices then fell 49% in the 1930s (Fishback et al., 2010), and 20-25% of mortgages went into default. The Home Owners’ Loan Corporation (HOLC) was established in 1933 to deal with the large number of mortgages in default during the Great Depression. It provided loans at 80% of appraised value for homes worth $20,000 or less ($336,000 today), and sought to keep people in their homes by providing assistance in collecting unemployment insurance, seeking paid work, and even finding tenants to help pay the mortgage (Harriss 1951, pp. 67f.). The government printed $2 billion of bonds ($33 billion in today’s dollars) to purchase mortgages and then refinanced them at low interest rates – around 1-2 percentage points below market rates at the time. Second liens could be refinanced as well as overdue property taxes, but the total loan could not exceed 100% of assessed value. Anyone who sold their mortgages to the government got paid from the revenues of these bonds.

Nonetheless, there are a number of important differences between the housing problems in the 1920s and the housing problems today. Unlike the 1920s, home prices have increased quickly since the Great Recession. When the housing market peaked in July 2006 the S&P/Case-Schiller US National Home Price Index was 184.62; it bottomed out in February 2012 at 134. As of August 2017 the index stood at a new high of 195.05 (FRED, 2017).

But this does not mean that housing problems have disappeared. Housing is important because the amounts involved are enormous. A $30,000 car loan and $50,000 in education loans pale in comparison with $300,000 in mortgage debt. Greater loan amounts mean that interest and principal payments will take up a larger fraction of total household income and will have a greater impact on consumer spending for other household expenses. It also means more danger for those holding the Old Maid card, or the debt, if things go bad.
Housing is also a problem because the other types of debt (college loans, credit cards, motor vehicles, etc.) are rising while at the same time household income is stagnating and even falling when household size is taken into account. This leaves less income to make mortgage payments, thereby increasing the chance of mortgage default.

Too many homeowners remain underwater on their mortgages. As of March 31, 2017, 10.4% of homes had negative equity (Zillow, 2017). As we argued previously (Scott and Pressman, 2017), it is not just underwater mortgages that are an issue. There is also a problem for households with only a small amount of equity in their home. While technically they are not underwater, they are still struggling since they cannot sell their home and move somewhere cheaper because selling costs and moving costs (including a security deposit and one month’s rent on a new place to live) exceed their home equity and savings balances.

Various programs, enacted during the Obama administration, sought to help homeowners remain in their homes. The Home Affordable Modification Program (HAMP) put $75 billion on the table to refinance mortgages at lower rates, although all the money was not spent on helping homeowners. Fannie Mae and Freddie Mac were bailed out to the tune of $187 billion, so they could keep purchasing conforming mortgages and prop up the housing market. Still, many homeowners remain underwater with their mortgages and loans that were modified right after the start of the Great Recession are experiencing problems (SIGTARP, 2017). And, unlike the 1930s, HAMP support is temporary, with interest rates on loans already beginning to reset at higher levels. The future of Fannie and Freddie also remains uncertain under a Trump administration and Republican-controlled Congress.

This means that many homeowners, who continued paying their mortgage, face rising mortgage payments. Furthermore, the Federal Reserve has been raising interest rates and will likely continue to do so; those with variable rate mortgages (taken to keep monthly payments down) will see their mortgages reset at higher rates soon.

Scott and Pressman (2017) estimated that over 16% of homeowners in 2013 had less than 10% equity in their home. Half of this, or 8%, were homeowners with zero or negative equity in their homes. In the 2016 SCF, 6.3% of households have zero or negative equity and an additional 6.2% have less than 10% equity in their homes. These figures are taken after more than 4 million people lost their homes due to foreclosure and after home prices have appreciated above their previous peak in 2006. A decline in home prices will likely be devastating to these home owners, and they are likely to result in another round of foreclosures.

The problem is not just homeowners underwater or nearly so; there is also the problem of huge mortgage debt. According to the SCF, in the early 1980s home debt was 30% of home values. Homeowners, on average, owned 70% of their home. By 2006 home debt grew to 50% of home values despite the large appreciation of home values over this time period. In the 2016 SCF, home debt is 56.7% of home values. The combination of record foreclosures and record high home prices has not drastically raised Americans’ equity share in their homes; rather, it has continued to decline over time.

Home prices are rising much faster than the incomes needed to support them. As a result, more households rely on increasing levels of debt to finance homes. Even before the housing bubble many homeowners faced great financial difficulty as a result of homeownership. One standard rule of thumb is that mortgages should not exceed 2.5 times one’s gross income (up
from the previous rule of 2 times gross income). A median income household (making $56,515) can then afford a mortgage of $141,540. With 10% down, they can afford a $156,000 home; with 20% down, they can afford a $178,000 home. Even rounding up to $200,000 the problems are clear. According to FRED, in the Q2 of 2017, the median price of a US home was $317,200, more than 50% higher than what a family with a median income could afford. Looking at this same problem using the SCF data, we find that outstanding mortgage debt is over 3.5 times median adjusted incomes, more than one-third greater than the standard rule of thumb. This is why many households with middle-class incomes are regarded as house poor or nearly house poor. Furthermore, house prices have not increased homogenously throughout the US. Homes in low-income and middle-income zip codes experienced less home value gains compared to high-income zip codes (Joint Center for Housing Studies, 2017).

We can look at this problem in yet another way. According to Fannie Mae, monthly housing expenses (on principal, insurance, interest and property taxes) should not exceed 25% to 28% of gross income. A family with a median income of $56,516 could afford to pay $14,129-15,824 in housing-related expenses. With average mortgage interest rates at nearly 4% in August 2017, putting down 10% on a median-priced home results in monthly mortgage payments of $1,368, or $16,410 for one year. Already the recommended ratio has been exceeded. Adding another $1,200 for PMI and another $1,685 for the median annual property tax in the median state (Georgia), brings the total annual cost of home ownership to over $19,000. Even putting 20% down (which also saves PMI), the median household still cannot afford the median house. Anyway you cut it, a median household is unable to afford a median priced home.

Figure 3 Real median United States incomes and the Case-Shiller US Home Price Index, 1984-2016

![Real Median Income and Case-Shiller Home Price Index](image)

Source: FRED (2017)

The housing situation is even worse when we consider that many households were foreclosed on during the financial crisis. This lops off the bottom of the distribution the same way that multi-generational living cuts off the bottom of the income distribution and pushes up
median incomes, making the income declines of a typical household worse than the reported data shows.

Homeownership rates fell by 5.4 percentage points between 2004 and 2016 (from 69.1% in 2004 to 63.7% in 2016) (Bricker et al., 2017). Some of this was due to declining ownership of speculative or vacation properties. Nonetheless, the remaining homeowners should be those in the best financial circumstances; still they remain in precarious shape. Analyzing the SCF data we found that in the past ten years, 3.5% of households had a foreclosure. This percent equates to over 4.4 million American households going into foreclosure. CoreLogic (2017) estimates that from 2007 to 2016 over 7.5 million residential homes were lost due to foreclosure – see Figure 4 below. For comparison sake, according to the SCF between 1971 and 2006, 2.48% of homes were foreclosed; and this is for a 35-year period rather than a ten-year period. The Great Recession resulted in more home foreclosures and a steeper drop in home values than the Great Depression – even taking into consideration the larger population size (Zillow, 2011). According to William Hedberg and John Krainer (2012) only 10% of homeowners that had a foreclosure (or serious delinquency) were able to get a mortgage within the following ten years.

The combined fall in homeownership and the large number of foreclosures had two important effects. First, households experienced a tremendous loss of wealth. Second, the “surviving” homeowners of the Great Recession (a) are more likely in better financial shape because they were able to avoid foreclosure and (b) were able to rebound from the low housing market to experience the tremendous appreciation in housing values that occurred after 2012 (see Figure 3 above). Yet, almost 13% of homeowners still have less than 10% equity in their homes.

**Figure 4** Foreclosures of residential homes in the United States, 2007-2016

Source: CoreLogic (2017).

Finally, the large number of foreclosures, in conjunction with rising debt (mortgage debt plus consumer debt), decimated household wealth in the US. As is fairly well-known, the largest
asset by far for most middle-class families is the value of their home. Using SCF data we find that, for households between the 40th and 80th percentile, real net worth fell from a median of $339,000 in 2007 to $249,000 in 2016 – a reduction of 26.5%. In contrast, households in the top 10% saw their net worth increase 26.4% over the same period ($1.297 million to $1.64 million) (also see Wolff, 2016).

3.c) Financial Institutions and Finance

The precarious state of many US financial institutions is one reason that the Great Recession was so great. In the 2000s these institutions had too little capital (Mian and Sufi, 2014) and too many non-performing assets on their balance sheet. Had they been forced to report their loans at market value in 2008 or 2009, many large banks would have been bankrupt. Since then bank assets have improved (due to increasing home prices, incomes and mortgage refinancing). Bank capital also increased as a result of this, as well as capital coming from the Federal Reserve in conjunction with dividend-payment restrictions for institutions taking central bank capital.

Recent Federal Reserve stress tests of banks provide some additional good news. Simulating the impact of a recession on bank balance sheets, the Board of Governors of the Federal Reserve (2017a) found that the 34 largest banks in the US could all survive a recession; they would not go under as Lehman Brothers did in 2008. Nonetheless, we need to approach these tests with some degree of skepticism. One issue is how much are these results like extra security at airports, designed to assure the public that flying is safe rather than thwarting terrorists. Similarly, stress tests can be seen as a way to reassure the public that it is safe to bank while not really making the banking system any safer. The Financial Stability Board (2013) notes progress made to increase bank customer protections and oversight through the Dodd-Frank Act – especially regarding the swaps market.3 Yet, they also found financial regulations convoluted in their overall supervision and mitigation of systemic risks.

Thun (2012) has criticized stress tests as being far too conservative when estimating the possible risks facing banks, for underestimating bank linkages, and for ignoring correlations between the prices of different assets. Let us take these points up in turn. Banks may be able to survive a normal recession, such as what the US experienced between the end of World War II and the early 2000s, but not a great slump. And systemic risk is not accounted for in these stress tests, as was true before the Great Recession. Another concern is that stress tests focus on common and known risks, giving banks an incentive to hide their risky endeavors in ways that the stress tests will not detect. Even if these tests were sufficiently tough, and even with bank capital up to 10 percent, a small rise in bankruptcies could push capital below the point where banks can make more loans due to capital requirement constraints. So it is not clear that the problems facing financial institutions have been solved.

Perhaps our greatest concern at present is that Congress is pressuring the Federal Reserve to reduce regulations on financial institutions. At the same time, the Trump administration is opposed to regulations and is not disposed to enforce existing regulations—in the hope that banks will lend more and make riskier loans that can be hidden from regulators who are predisposed to look the other way. Even worse, the Trump administration and Congress continue

---

3 Similar things are true in the UK, where the Financial Stability Board claimed that the financial system had become more stable. They are likely true as well in other countries.
to talk about repealing the meager regulations that exist as a result of Dodd-Frank (Puzzanghera, 2017).

While mortgage-backed securities, and their offshoots, such as CDOs and the ARSs mentioned at the beginning of this paper, have been identified as a main cause of the 2008-9 financial crisis (Jarrow, 2011; Stiglitz, 2009), today we have their close cousins to deal with – collateral loan obligations (CLOs). CLOs are bundles of risky loans that have been packaged together in order to get a high credit rating from Moody’s and Standard & Poor’s (based on their rating algorithms). These packages typically include student debt and auto title loans—two things discussed earlier in this paper. Many individuals and institutions buying CLOs believe that the default risk is near zero, just as buyers of mortgage-backed securities and CDOs believed that defaults on home loans were not possible following the 1993 Supreme Court decision that prohibited homeowners from reducing their mortgage debt in bankruptcy (Taub, 2014). Dodd-Frank was supposed to stop this process of securitization and sale, but the SEC has allowed it to continue. This puts on the balance sheets of financial institutions, as well as households, assets that are AAA rated and that many believe are default proof. Their owners seek high yields in an era of low interest rates, and believe that there is little risk in holding these assets because there cannot be a large asset price decline or market crash, and because people need their cars and cannot default on auto-title loans or student debt.

One final problem looming on the horizon concerns tax reform. As of this writing there is talk of both reducing the mortgage deduction (possibly indirectly, through a large increase in the standard deduction), as well as reducing or eliminating the deduction for property taxes. These actions will hurt many middle-income families owning homes. Some families are able to pay their mortgage only because their property taxes and mortgage interest are tax deductible; they may no longer be able to continue paying their mortgage if tax benefits from homeownership are reduced but are not cut for middle-income households (as is currently being proposed). Eliminating these deductions will also put downward pressure on home prices, worsening the problem of underwater and near-underwater mortgages, as well as the wealth of middle-income households.

4. Conclusion – preventing another Great Recession

Minsky (1982; 1984) argued that financial crises would be somewhat infrequent because memories remain vivid in the post-crisis period and lenders turn conservative. On this point Minsky was wrong. First, as we saw above, the US economic history has been a history of frequent economic crises – on average, one every 8-9 years. This indicates that memories are relatively short rather than long. Minsky also underestimated the role of greed as part of human nature and the desire to be a little better than everyone else. In contrast, Veblen [1899] understood this point well, as do Frank and Cook (1995). In his recent book, The Broken Ladder, Keith Payne (2017) shows this to be part of our evolutionary nature and our psychological makeup. He cites many studies demonstrating that humans, and our evolutionary ancestors in the animal world, care about relative position. One of the most noteworthy involves monkeys who were given either grapes (which they love) or cucumbers (which they merely like) for returning a stone. Monkeys that were happy getting cucumbers got angry when other monkeys received a grape for performing the exact same task. They frequently threw their cucumber at the experimenter; and they stopped returning stones. The importance of relative incomes, as modeled by James Duesenberry (1949), leads people to consume based not on their own incomes, but relative to those within the culture where
they live and want to be associated. In addition, Thorstein Veblen’s (1899) pecuniary
emulation takes Duesenberry’s insights a step further, showing that people want to consume
one social strata above what their current income allows. Eventually, these cultural and
psychological tendencies can create problems for people in the form of excessive debt, and
for financial institutions that make loans to households that are overwhelmed by debt and
unable to repay their debt obligations, especially in an era of stagnant or declining real
incomes.

There are many reasons to be concerned about the US recovery from the Great Recession.
This paper argues that the US economy has not escaped from the problems that created the
Great Recession. Real median household income, adjusted for household size, has fallen
3.3% over the past decade. Larger interest payments on past debt has reduced the living
standard of the median household another 3.1%. From business-cycle peak to now
(presumably close to another peak), the standard of living for the median US household
(adjusted for household size and consumer debt interest) is 6.4% below the level of 2007.
Further, household debt levels remain high, especially relative to current household income
levels; and housing remains a problem. Many homeowners are underwater or nearly
underwater on their mortgages; and large financial institutions still face inadequate regulation.

The good news is that we can mitigate these problems and reduce the chances of another
Great Recession.

We could raise taxes on the very rich. Thomas Piketty (2014; Piketty et al., 2014) cites low
top marginal tax rates as one reason CEO pay has soared. It also explains why CEOs have
sought to cut labor costs dramatically and why wages have not kept up with either productivity
growth or economic growth. Tax hikes on the richest Americans will counter this, and will help
to prop up wages. The additional government revenues could finance more generous
programs to help households during hard economic times, making them less dependent on
high-interest loans in times of personal financial crisis. Liberalizing bankruptcy laws would
help, making it easier for households to escape from suffocating debt.

Greater restrictions can be placed on financial institutions, which are much larger today than
ten years ago. In addition, regulations will require real teeth and substantial penalties for
those who break the rules. Returning to Glass-Steagall should also be a top priority. Individual
consumers cannot be expected to battle a few large financial institutions on their own. The
Dodd-Frank financial reform bill established the Consumer Financial Protection Bureau
(CFPB) to curb egregious practices by financial institutions.4 This is not enough.

Third, housing costs need to be brought under control so that debt ratios can become more
manageable if incomes rise. The tax deductibility of mortgage interest and property taxes
adds to price pressures. It also benefits wealthier households in higher tax brackets who can
afford more housing because of the tax benefits given to them. This tends to have a
cascading effect, pushing up the price of other homes, even though their owners will get few
tax breaks. One simple reform would be to convert the tax deduction for mortgage interest
and property taxes into a refundable tax credit so that everyone benefits to the same extent
from this tax provision. Another way to control housing prices would be to provide incentives
to increase the housing stock, especially in those areas where prices have risen rapidly

4 Even this minor and weak reform was fought by banks, and Republicans refused to approve anyone
nominated to head the agency for a long time.
during the past few decades. One solution might be to provide incentives to convert abandoned shopping malls into housing units, similar to what has been done in some places with abandoned factories. Of course, given the large percentage of underwater and near-underwater mortgages, these changes must be made very slowly.

Our big fear is that it is unlikely that such policy changes will be made in time to avoid another great crisis. However, if they are not made, the next recession seems destined to result in massive bankruptcies and layoffs. Homeowners will again lose their homes. Households holding CLOs will find their assets are no longer liquid. In sum, it will appear that we really never escaped the Great Recession; rather, we papered over our problems so that they lay dormant, waiting to create havoc for the US economy.

References


Author contact: pressman@monmouth.edu and rscott@monmouth.edu

SUGGESTED CITATION:

You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-83/
Abstract
Since Thomas Piketty’s work took the economics discipline by storm in 2014, the study of economic inequality has quickly moved to the center of academic inquiry for the first time since the nineteenth century. But why was the topic of inequality and distribution largely neglected by mainstream economists for much of the twentieth century? In examining key economic treatises and textbooks, this paper argues that it was the hegemonic rise of neoclassical economics which effectively marginalized the issue of economic distribution in the twentieth century. It contends that three central theoretical pillars were most responsible for this development: marginal productivity, a utility theory and Pareto optimality.

Introduction
Six years after the worst economic depression in decades, a 600-page book on economic inequality took the United States by storm, becoming a surprise best seller. Despite its populist message, the book was largely in conversation with orthodox economic thinkers, not radical ones, and thus retained the former’s conceptual categories, worldview and language. And yet, despite its conventional approach – or perhaps because of it – the book sent a shockwave through the field of economics. Using the master’s tools, it took a hammer to some of the most cherished tenets of contemporary economic thought, disrupting accepted wisdoms, destabilizing basic convictions and upending theoretical models.

The source of the book’s power lay in the fact that it refused to countenance the notion that the study of economics could and should be insulated from distributional questions. While many economists of the era peddled feel-good stories about the harmonious interests of workers and elites, this book politicized economic thought by revealing that capitalism produced winners and losers, productivity and inequality, progress and poverty.

Inspiring grassroots organizations, social movements and various reform proposals, the book became not only an American or even European phenomenon, but a global one. Tedious, unwieldy and redundant at times, it nevertheless managed to capture the imagination of a shell-shocked generation still reeling from a major economic crisis. It did so for two main reasons. First, it offered a simple argument: Inequality was caused by the ever-increasing concentration of unearned wealth in the hands of an elite class of unproductive rentiers. Second, it offered a simple solution: A single tax on wealth that would prevent such elites from profiting off the mere possession of property. With such a straightforward message, small wonder the book became an international sensation.

The dismal science had given birth to an economic rock star. His name was Henry George. The year was 1879.1

The Piketty effect

A few years ago, I opened my review of Thomas Piketty’s *Capital in the 21st Century* in the *Raritan Quarterly Review* with this “bait and switch” vignette. I thought the striking similarities between George and Piketty revealed that while history does not repeat itself, the “Pikettymania” that washed over the world in 2014 might bring forth once more an era in which – much like during the “Gilded Age” of Henry George – economic inequality was at the forefront not only of economic thought but political agitation, social anxiety and cultural discourse.²

Looking back now to those heady days in 2014, it is clear that Piketty’s groundbreaking study was just the beginning. The floodgates of inequality studies have been opened. The wave ushered in by Piketty has, in the past few years, come in many shapes and sizes: We now have global analyses such as Branko Milanovich’s *Global Inequality*, centuries-long histories such as *Unequal Gains*, and a collected volume dedicated entirely to the economic agenda titled *After Piketty*. The dramatic titles of other recent books reveal the current mood of inquiry, be it Thomas Shapiro’s *Toxic Inequality: How America’s Wealth Gap Destroys Mobility, Deepens the Racial Divide, & Threatens Our Future*, Dean Baker’s *Rigged: How Globalization and the Rules of the Modern Economy Were Structured to Make the Rich Richer* Steven Teles and Lindsay Brink’s *The Captured Economy: How the Powerful Enrich Themselves, Slow Down Growth, and Increase Inequality* or Brian Alexander’s *Glass House: The 1% Economy and the Shattering of the All-American Town*. It appears, that the “1 percent” have not only been gobbling up much of the wealth and income these past few decades but, in recent years, also the attention of economists, journalists and public intellectuals.³

The “Piketty effect” has spread into political and policymaking circles as well. If there is one constant in the left rhetoric of Bernie Sanders, the most popular politician in the United States in 2017, it is his dogged emphasis on the massive wealth disparities between the super-rich and the “99%”. A visit to inequality.org reveals, moreover, a long list of think tanks, academic centers and public interest groups who now focus on inequality, be it the Economic Policy Institute, the Washington Center for Equitable Growth, or the LSE International Inequality Institute. Inequality has even seeped into the staid world of central banking. U.S. Federal Reserve Chair Janet Yellen spoke at a Fed conference in Boston in the fall of 2014, just as Piketty’s book was taking off. “The extent of and continuing increase in inequality in the United States greatly concern me,” Yellen said. “I think it is appropriate to ask whether this trend is compatible with values rooted in our nation’s history, among them the high value Americans have traditionally placed on equality of opportunity.” As the *New York Times* rightly

---

noted at the time, “by the cautious standards of central bankers,” Yellen’s words were “downright radical.” As we will see, this is not how Fed Chairs spoke about inequality before Thomas Piketty.4

The goal of this article, however, is not to focus on the rebirth of inequality economics. Rather, it is to try to answer an oft-overlooked question: what took so long? For all the attention it has garnered, it is easy to forget that Piketty’s book became a smash hit not because of its explanatory power (few have actually agreed with his \( r > g \) model) but rather mostly thanks to his fairly straightforward empirical project which measured, over the course of the twentieth century, how income and wealth were distributed in the United States and Western Europe. While Piketty’s impressively Sisyphean archival work—especially with government tax records—should be commended, the question still must be asked: If western society in the mid-to-late twentieth century was hardly lacking in economists, how come Piketty’s study had not already been carried out on numerous occasions? In the few cases in which similar, albeit far more modest, studies were undertaken, why were they mostly neglected? Or, to put it another way: If Henry George’s Progress and Poverty and the question of economic inequality took the Western world by storm in the nineteenth century, why was it marginalized for much of the twentieth, only to return with a vengeance in the twenty-first?5

I am not the first person to recognize the fall of distributive economics in twentieth century Western thought. In April 1996, Sir Anthony Atkinson—who Piketty has labeled the “godfather” of inequality economics—gave the presidential address to the Royal Economic Society of England. Atkinson opened his talk by noting that “the subject of income distribution has in the past been marginalized. For much of this century, it has been very much out in the cold.” Continuing, Atkinson noted how, for the past 50 years, only about 4 percent of The Economic Journal articles, one of the leading economics journals for much of the twentieth century, had dealt with income distribution. In comparison, Atkinson demonstrated how international economics produced on average four times as many articles during that time span. In the 1970s, just as income inequality was beginning to significantly rise, a quarter of the articles in the Journal were on globalization but only about one article per a year was on distribution or inequality.6

Why did this happen? Relatively low income inequality in the middle decades of the twentieth century certainly played a role in its diminished position, especially during the post war years. When economist Simon Kuznets invented the “Kuznets Curve” theory in 1954, which posited

5 For the typical enthusiastic response to Piketty’s data yet disagreement with \( r > g \) see Paul Krugman, “Why We’re in a New Gilded Age,” New York Review of Books, May 8th, 2014; James K. Galbraith, Kapital for the Twenty-First Century? Dissent, (Spring, 2014); To be sure, there are a number of important exceptions in which economists did study inequality in the twentieth century. See the life’s work of Anthony Atkinson, including such works as The Economics of Inequality (New York: Oxford University Press, 1983); Claudia Goldin and Robert Margo, “The Great Compression: The Wage Structure in the United States at Mid-Century,” NBER Working Paper 3817 (August, 1991); James D. Smith, eds., Modeling the Distribution and Intergenerational Transmission of Wealth (Chicago: Chicago University Press, 1980).
that as capitalist societies develop their inequality tends to go down, he was basing his hypothesis on empirical data which showed a decline in inequality since the stock market crash of 1929. This allowed future economists, when confronted with the issue of inequality, to mumble a few words about the Kuznets curve and move on to what they felt were more pressing inquiries. As late as 1980, a leading economist could still brush off the issue of inequality by noting that “income inequality was just about the same in 1977… as it was in 1947.” Some economists of the era even believed that labor’s share of income was so consistently stable in the postwar years that economic distribution must be the workings of some natural law. 

These explanations for the marginalization of inequality, however, will not suffice. When actual income inequality began to climb in the 1980s, the study of income inequality did not follow suit. What is more, the neglect of inequality economics in the twentieth century is so striking it cannot be explained away so easily. Economists did not start ignoring inflation or unemployment in eras when they happened to be low. What then was the main impetus? This paper argues that it was the hegemonic rise of neoclassical economics which effectively marginalized the issue of economic distribution in the twentieth century in favor of the maximization of economic production. More specifically, I contend that three central theoretical pillars of neoclassical economics were most responsible for the downplaying of inequality and distribution: marginal productivity, a utility theory of value and Pareto optimality.

While some economists frame the history of economic thought as being driven mostly by the internal improvement of the discipline in its long march towards scientific truth, these neoclassical pillars did not emerge in an historical vacuum strictly because they were empirically more accurate than past models. Historians in recent years have shown that neoclassical economics was shaped by an assortment of political, social and cultural forces be it the rise of consumer culture, corporate finance, modern psychology, social democracy or thermodynamic physics. In this brief article I cannot touch on all the forces that led to the rise of neoclassical economics. I will, however, stress one crucial and oft-overlooked engine of neoclassical theory that is most relevant to our discussion: the desire to downplay, marginalize and mitigate distributional questions and conflicts because they were deemed either too dangerous, moralizing or unimportant. In other words, the meteoric rise of neoclassical economics did not only lead to a sharp decline in distributive economics, but was partially constituted by this very goal.

---


The centrality of distribution to classical economics

Henry George’s *Progress and Poverty* was the last great work of what historians have referred to as “classical” economics. At the heart of his book was a basic question regarding the relationship between economic growth and economic distribution. “Where the conditions to which material progress everywhere tends are most fully realized,” George pointed out in his opening statement, “we find the deepest poverty, the sharpest struggle for existence, and the most enforced idleness.” Continuing, he stunned many readers by declaring that “material progress does not merely fail to relieve poverty—it actually produces it.”9

It was this notion that capitalist development brought with it great wealth to some but terrible suffering to others that gave George’s book its radical and critical bite. Yet despite this – and the fact that his call for a “single tax” on the unearned rent of land monopolizers directly challenged the most basic liberal tenets of private property – his work was still very much in line with the English classical tradition of Adam Smith, David Ricardo and John Stuart Mill. While not reaching the same subversive conclusions as George, these men had also been most interested in understanding how the fruits of market production were divided between the three classes of society – land-holding aristocrats, profit-seeking capitalists and wage-laboring workers. They did so in part because they believed distribution to be an important social and moral issue that should not be ignored. But they also did so because they believed that it was the social relationships between these three social classes (rather than mere supply and demand) that determined not only the rate of compensation of each class in the form of rent, profit and wages but the price of all market commodities. In classical economics, in fact, the market does not really set the price of goods at all. Rather, the “natural price” of any commodity is set by the rates of wages, rent and profit which, in turn, are set by the social relations between workers, capitalists and landholders. In this theoretical world, which focuses mostly on economic production, exchange serves only as the tool through which market prices become aligned with natural prices. In short, to study any aspect of “the economy” in classical economics, you had to study the distribution of wealth and income.10

George’s focus on inequality, therefore, was no great departure from the classical economists who came before him, especially Ricardo and Mill. They too, had placed the distribution of income at the center of their discipline. Ricardo, for instance, famously began his magnum opus of 1817 by stating:

“The produce of the earth - all that is derived from its surface by the united application of labor machinery and capital, is divided among the three classes of the community: namely the proprietor of the land, the owner of the stock of capital for its cultivation, and the laborers by whose industry it is cultivated. But in different stages of society, the proportions of the whole produce of the earth which will be allotted to each of these classes, under the names of rent, profit and wages, will be essentially different...to determine the laws which regulate this distribution, is the principal problem in political economy.”

Mill would continue Ricardo’s emphasis on distribution, noting how “it is only in the backwards countries of the world that increased production is still an important object; in those most

---

9 Henry George, *Progress and Poverty: An Inquiry into the Cause of Industrial Depression and of Increase of Want with Increase of Wealth* (New York: 1879), 5.
advanced, what is economically needed is a better distribution.” 11

George, moreover, was not only continuing the tradition of classical economics but also that of American republicanism and producerism. The United States came into being as a society of freehold farmers who placed an enormous emphasis on the basic freedom (if you were white) to receive a “full return of one’s labor.” Steeped in the labor theory of value not of Karl Marx but Benjamin Franklin, nineteenth century white men in the United States were raised to believe that an unequal distribution of wealth or income was far from natural and therefore must stem from exploitative social relations in which laborers do not receive all that they have produced. As the prototypical American economic thinker Edward Kellogg noted in his 1849 book Labor and Other Capital, “to obtain labor without rendering a fair equivalent is also a violation of the rights of property.” In his eyes, like that of most Americans, this meant that “the great disparity in the conditions of the rich and poor is the natural result of unjust laws.”12

Such Americans also found a basis for their claims in European classical economics. Following in the footsteps of Ricardo and Mill, most nineteenth century economic thinkers in Europe and the United States believed that a capitalists’ profit stemmed from his selling of goods for more than his workers had been paid to make them. This approach positioned laborers and capitalists in a zero-sum struggle for the economic surplus. “If... wages should rise,” Ricardo repeatedly stated, “profits would necessarily fall.” Or, as Mill noted in 1869, if a capitalist “has to pay more for labour, the additional payment comes of his own income.”13

There were, of course, plenty of conservative economists who pushed back on this idea throughout the nineteenth century. Economists like Frederic Bastiat and Nassau Senor argued that the profits of capital did not come from their power struggle with labor or their appropriation of the surplus but rather from risk, abstinence, skill, entrepreneurship and other positive qualities. “Capital has its roots in three attributes of man,” Bastiat typically declared in 1850, “foresight, intelligence, and thrift.” Yet try as these conservatives might to separate profits from wages, the labor theory of value which stood at the center of classical economics made this very hard to do since it was assumed that wealth was created mostly by workers. To make matters worse for such economists, by the late nineteenth century a far more radical thinker than Ricardo, Mill or even George had turned to the classical labor theory of value in Europe in order to argue that the exploitative basis of capitalist accumulation meant it must be overthrown. All across Europe, socialists turned to Karl Marx’s economic writings in order to prove that capital profits were nothing more than the appropriated “surplus value” of exploited labor. It was, in part, in the midst of these political pressures that our first neoclassical pillar was born.14

---


Marginal productivity and the justice of capitalism

The year was 1899 and Columbia Economics Professor John Bates Clark - the undisputed American father of neoclassical economics – was gravely concerned. He felt that all this talk in Europe and the United States about inequality and exploitation was threatening to destabilize the very foundations of a modern, capitalist society. As he saw it,

"the welfare of the laboring classes depends on whether they get much or little; but their attitude toward other classes—and, therefore, the stability of the social state—depends chiefly on the question, whether the amount that they get, be it large or small, is what they produce. If they create a small amount of wealth and get the whole of it, they may not seek to revolutionize society; but if it were to appear that they produce an ample amount and get only a part of it, many of them would become revolutionists, and all would have the right to do so. The indictment that hangs over society is that of 'exploiting labor.' 'Workmen' it is said, 'are regularly robbed of what they produce. This is done within the forms of law, and by the natural working of competition.' If this charge were proved, every right-minded man should become a socialist; and his zeal in transforming the industrial system would then measure and express his sense of justice."\(^{15}\)

Luckily for “the stability of the social state,” in the same book in which Clark voiced these concerns, he also presented a novel economic theory which claimed to prove that the distribution of wealth in a competitive market society was, in fact, inherently just and that there simply was no such thing as labor exploitation. On the contrary, according to Clark, every class in society got what it deserved for it earned what it had produced. The book was titled *The Distribution of Wealth* and its main goal was made perfectly clear in its opening pages:

"It is the purpose of this work to show that the distribution of the income of society is controlled by a natural law, and that this law, if it worked without friction, would give to every agent of production the amount of wealth which that agent creates."\(^{16}\)

Unlike in classical economics, where laborers and capitalists fought over the same pool of surplus production, Clark sought to insulate the economic mechanism through which wages were determined from the economic mechanism through which profits and rents were determined. Here too, Clark was refreshingly open about the political reasons for wanting to do this, noting that “it was the claim advanced by Henry George… that first led me to seek a method by which the product of labor everywhere may be disentangled from the product of cooperating agents and separately identified.” As neoclassical economist Frank Fetter later recognized, “one can hardly fail to see on almost every page” of Clark’s writings the single-tax specter of Henry George.\(^{17}\)


\(^{16}\) Clark, *Distribution of Wealth*, vi.

Clark’s great innovation was to treat the labor of workers, the capital of capitalists and the land of landholders as three utterly separate “factors of production” whose respective incomes in the form of wages, profits and rents were determined in three utterly separate markets by their owner’s own marginal productivity. According to Clark, a worker earned 5 dollars an hour not because his boss may be exploiting him but rather because that was the contribution of his final (or marginal) hour of labor to the production process. On the other hand, a capitalist earns $5,000 in profit not because he had the social power that came with owning the means of production but rather because this was the productive contribution of the machinery he owned. The moral of the model was obvious: the distribution of wealth in free market societies was inherently fair. So long as the government or unions didn’t interfere in the workings of a competitive market, both worker and capitalist would receive their just deserts.18

As the title of his book makes plain, Clark clearly did not ignore or downplay the issue of economic distribution. Quite the opposite in fact. Yet as his marginal productivity theory grew to become one of the central pillars of neoclassical economics in the twentieth century, its effect was largely to sideline questions of distribution. For if each person in society received what they had produced, then what mattered most was not the issue of economic inequality but rather economic productivity. So long as neoclassical economists studied ways in which to increase productivity, they had little need to examine the mechanisms through which it was distributed. As a result, the neoclassical economists who followed in Clark’s footsteps put far less of an emphasis on distribution. For example, Yale Economics Professor and neoclassical savant Irving Fisher derided socialists who thought “the problem of economic mass welfare is primarily one of distribution,” arguing, rather, that “it is primary one of production.” Clark’s claim also took the ethical sting out of inequality. Since each person gets what he deserves, whatever inequality that does exist in society is legitimate. As noted by Cambridge economist Alfred Marshall, Clark’s avid follower in England and author of the leading neoclassical textbook of the first half of the twentieth century, “most earn just about what they are worth.”19

To see the long-term impact of Clark, look no further than Paul Samuelson’s Economics: An Introductory Analysis, which became the best-selling economic textbook of all time in the latter half of the twentieth century. In the seventh edition from 1967, there are over 800 pages. How many directly examine the issue of wealth or income inequality? About two dozen. Why? We will get to the other reasons in a moment but it is interesting to note that in his discussion on economic distribution Samuelson instructs his readers “to appreciate J.B. Clark’s advance over such classical economics as David Ricardo.” Moreover, he argues not only that “the Clark neoclassical theory of distribution, although simplified, is logically complete and a true picture of idealized competition,” but also that empirical evidence “seems to provide rough corroboration for [his] theories of production and marginal-products.”20

The Chicago School’s theory of “human capital,” which took off in the 1960s and 1970s, also reveals how marginal productivity led many economists to focus more on productivity than

---

distribution. In the late twentieth century, human capital theory quickly became the most dominant approach by labor economists for not only valuing people but explaining inequality. In treating people as capitalized factors of production much like machines, human capital theory posited – just like Clark – that labor wages were largely determined by labor productivity which, in turn, were largely determined by how much a worker “invested” in themselves to improve the “rate of return” on their human capital. In following Clark’s theory of marginal productivity, these economists usually did not bother to examine how the economic pie or national income was divided between labor and capital. Rather, since they assumed that each worker earned what he or she in fact produced, they focused only on the question of labor productivity and how it could be increased via self-investments in training and education. Just as Clark had intended over one hundred years before, gone were the classical economic questions regarding the ways in which social or power relations influenced the distribution of wealth or income between labor and capital.21

The utility theory of value and spiteful egalitarianism

Another key reason neoclassical economics tended to downplay distribution was its shift from a labor theory of value which focused on production to a utility theory of value which emphasized exchange and consumption. Unlike classical economics, in neoclassical economics the distribution of wealth between social classes plays no role in the determination of commodity prices. This is not only because there are no social classes in neoclassical economics, only individual utility-maximizing exchangers, (incredibly, labor can buy capital in neoclassical models just as capital buys labor) but also because neoclassicists believe it is the prices of goods that help determine the rate of wages, profit and rent and not vice versa. The price of any commodity, meanwhile, is determined by the final (or marginal) amount of subjective utility it offers individual consumers and has little to do with the distribution of wealth.22

The move from a producerist labor-theory-of-value to a consumerist utility value theory also marginalized distribution by offering an alternative meaning of freedom and wellbeing. While American farmers and European socialists had used a labor theory of value to argue that to be free and prosperous entailed a full return of the fruits of their labor, the invention of marginal utility reflected a consumerist turn which envisioned freedom as the consumption of the fruits of industrial progress. Laborers should not care what the profits of their employers were in comparison to their own wages, the argument went, so long as their “standard of living” and consumer comfort was increasing.23

A perfect example of this consumerist marginalizing of inequality took place in the late nineteenth century in the United States, just as neoclassical economics was coming into being. As a PhD Student at Columbia in the 1880s, Charles Barzilai Spahr wrote a


dissertation that likely would not have been written in the mid-twentieth century. Published later as a book titled *An Essay on the Present Distribution of Wealth in the United States*, Spahr meticulously mined the taxation data at his disposal to make one basic point: as time passed, the distribution of wealth in the United States was becoming more unequal. “Seven-eighths of the families [in America] hold but one-eighth of the national wealth,” Spahr concluded, “while one per cent of the families hold more than the remaining ninety-nine.” More than a century before Thomas Piketty, Spahr had discovered the infamous “one percent.”

Spahr, however, has been completely forgotten in large part because of the scathing review he received in the premier academic journal of the era by Columbia University Economics Professor Richmond Mayo-Smith, who just so happened to be a close colleague of John Bates Clark. “Having shown that property and incomes are unequally distributed and that (in his opinion) the inequality is increasing,” Mayo-Smith wrote, “Dr. Spahr seems to think that his task is ended. But that is only the beginning. The real question is whether such a concentration of wealth is not a good thing for the whole community.” Continuing, Mayo-Smith reflected the turn to subjective utility value by arguing that “the happiness of individuals is measured not according to their ownership of property ... but according to their command of the enjoyments of life.” Mayo-Smith’s argument that labor should focus on its subjective consumer enjoyments rather than the unequal gains of capital was repeated numerous times in the twentieth century. For instance, just as economic inequality was slowly returning to American economic discourse in the late 1990s, Harvard Professor and National Bureau of Economic Research President Martin Feldstein tried to shut it down. He did so by arguing that such a focus on inequality stemmed from an ideology of “spiteful egalitarianism” and that economists need not make such troublesome comparisons:

> “According to official statistics, the distribution of income has become increasingly unequal during the past two decades. A common reaction in the popular press, in political debate, and in academic discussions is to regard the increase in inequality as a problem that demands new redistributive policies. I disagree. I believe that inequality as such is not a problem and that it would be wrong to design policies to reduce it. What policy we should address is not inequality but poverty.”

Yet the relationship between the rise of a utility theory of value and the demise of inequality economics in the twentieth century was not nearly so cut and dry. In the first generation following its inception - and crucially before the “Paretian Revival” of the 1930s that would place “Pareto Optimality” at the very center of neoclassical economics – marginal utility actually led many economic thinkers to focus more on the relationship between economic inequality and efficiency – not less. In fact, in the late nineteenth century the socialist Fabian society and its leaders Sidney Webb and George Bernard Shaw became enthusiastic proponents of the utility theory of value and its potential for reinvigorating the case for economic equality. This is because they, and many other utility theorists in the early twentieth century, were very “Benthamite” in their approach to social welfare.

---

Like the eighteenth century utilitarian philosopher Jeremy Bentham, these economic thinkers believed that in order to measure the total welfare of a society all one had to do was sum up the utility consumed by every individual. They argued that it was theoretically possible to add together individuals’ subjective feelings of satisfaction or happiness because individual utilities were comparable between people and thus also capable of aggregation. This interpersonal comparison approach to utility theory led to radically egalitarian conclusions: Since all neoclassical economists believed in the principle of diminishing marginal utility, Benthamites logically concluded that the marginal utility of a rich man was lower than that of a poor man. This meant that even if one disregarded the moral elements of inequality, the most efficient way to maximize social welfare was to redistribute money from the rich to the poor. As Cambridge economist Arthur Pigou (hardly a radical socialist) explained in his renowned 1920 book *Economics of Welfare*, egalitarianism was the most efficient way to maximize welfare because “more intense wants to be satisfied at the expense of less intense wants must increase the aggregate sum of satisfaction.” Pigou, therefore, concluded that any redistributive policy “which increased the proportion of the national dividend by poor persons, provided that it does not lead to a contraction of the dividend…will, in general, increase economic welfare.”

**Pareto optimality and the dangers of comparison**

The egalitarian era of marginal utility theory, however, was short-lived. Even at its peak in the early 1900s, there were numerous leading neoclassical economists, such as Stanley Jevons, Francis Edgeworth and Vilfredo Pareto, who rejected such arguments for equality. The Italian Pareto led the charge in this regard. A classical liberal before he turned to Italian Fascism, men like Pareto were disturbed by the notion that marginal utility theory could be used as a tool to legitimize the redistribution of wealth from rich to the poor. To counter these arguments, he made key modifications to utility theory in his 1906 book *Manual of Political Economy* that disabled such egalitarian arguments while also marginalizing the issue of inequality. First off, Pareto argued that one could not make interpersonal comparisons of utility. Since utility was subjective desire, Pareto argued, it was simply impossible to compare one person’s marginal utility – no matter how rich they happened to be – with that of another. This modification to utility theory led Pareto and other neoclassicists to claim that “Benthamite” economists could not compare the utility of two people nor could they measure social welfare by adding up the utility of all individuals in a given society.

Basing his analysis on these key assumptions, Pareto came up with his own definition of social optimality. Since the utilities of individuals could not be compared or aggregated, Pareto argued, it was not necessarily economically optimal to take from the rich and give to the poor because it would not be clear if this was a net utility gain for society or not. In a world where interpersonal utility comparisons could not be made, Pareto continued, a definite efficiency improvement could only take place if one person was made better off without injuring anyone else – even in the slightest. According to Pareto’s logic, even though a starving man could use a dollar far more than a millionaire, if that millionaire felt even an

---

Inkling of pain in giving up that dollar, then by definition redistributing the money was inefficient and not optimal. The principle of “Pareto Optimality” which still dominates neoclassical economics was born. With its rise, economists’ interest in inequality would significantly wane.  

In time, Pareto optimality would form the core of neoclassical economics’ case for free-market liberalism. As Pareto (and his Lausanne University predecessor Leon Walras) had argued, the wonder of the unregulated free market lay in its ability to always lead society to a Pareto optimal point in which no more transactions could be made that improved one person’s lot without harming anyone else’s. Translated into highly mathematical terms, the free market’s almost magical ability to result in Pareto optimal allocations came to be known as “the first fundamental welfare theorem of welfare economics.” For the generations of neoclassical economic students who followed, this theorem would be key, not only because proving it required a high-level of mathematical expertise, but also because it gave a scientific veneer to the idea of “the invisible hand.” As Paul Samuelson would explain to generations of economics students in his 1960s textbook, the idea of Pareto optimality had shown that

“Adam Smith, in his talk about an Invisible Hand, which led the selfish actions of individuals toward so harmonious a final result, did have some point...Under perfectly perfect competition...where the genuine desires and well-being of individuals are represented by their marginal utilities...then the resulting equilibrium has the efficiency property that ‘you can’t make any one man better off without hurting some other man.’ What does this mean exactly? It means that a planner could not come along with a slide rule and find a solution, different from the laissez-faire one, which could improve the welfare of everyone.”

The idea of Pareto optimality not only helped to legitimate free markets but it also reflected the bourgeois ideology of classical, Lockean liberalism: the ownership of private property was a natural right above any artificial state intervention. As such, no infringement on private property, no matter how large the wealth disparities in a society may be, could possibly be socially desirable. Implicit in this argument was the claim that economists need not focus on inequality since any attempts at redistribution would distort the workings of the free market and thus lead to sub-optimal allocation points.

Pareto’s approach to utility theory did not catch on right away. In fact, it was “Pigouvian” and not “Paretian” welfare economics that seemed to be more popular in the early twentieth century. As a result, one can still find in the early twentieth century many studies on wealth and income inequality, including a groundbreaking report by the American National Bureau of Economic Research in 1920, which rejected many of Pareto’s theories. All this changed, however, in the 1930s and 1940s. Within the span of less than two decades, neoclassical economics swung completely to the side of Pareto Optimality. While the Paretian Revival would encompass the entire discipline – from socialists like Oskar Lange to liberals like Paul

---

Samuelson to conservatives like Milton Friedman – the man most responsible for bringing this change about was Lord Lionel Robbins, a London School of Economics Professor and member of the “neoliberal” Mont Pelerin Society who gave Austrian School economist Friedrich Hayek his first job in England.\(^{32}\)

In the late 1930s, Robbins re-iterated Pareto’s arguments against the ability to make interpersonal utility and, therefore, comparisons between the rich and poor. In so doing, however, Robbins went one step further than Pareto by reaching the conclusion that any analysis of inequality or distribution was inherently normative and, therefore, should play a limited role, if any, in the positivist science of economists. Robbins’ 1938 article on interpersonal utility comparisons is widely regarded for turning the tide in the neoclassical economic approach to distribution. Here is the most quoted section:

“But as time went on, things occurred which began to shake my belief in the existence between so complete a continuity between politics and economic analysis…. I am not clear how these doubts first suggested themselves; but I well remember how they were brought to a head by my reading somewhere – I think in the work of Sir Henry Maine – the story of how an Indian official had attempted to explain to a high-caste Brahmin the sanctions of the Benthamite system. ‘But that,’ said the Brahmin, ‘cannot possibly be right – I am ten times as capable of happiness as that untouchable over there.’ I had no sympathy with the Brahmin. But I could not escape the conviction that, if I chose to regard men as equally capable of satisfaction and he to regard them as differing according to a hierarchical schedule, the difference between us was not one which could be resolved by the same methods of demonstration as were available in other fields of social judgment.”

In conclusion, Robbins declared that “I still cannot believe that it is helpful to speak as if interpersonal comparisons of utility rest on scientific foundations – that is upon observation and introspection…I still think, when I make interpersonal comparisons, that my judgments are more like judgments of value than judgments of verifiable fact.”\(^{33}\)

Not everyone agreed with this argument. In the same year as Robbins article was published, Sir Roy Harrod warned that “if the incomparability of utility to different individuals is strictly pressed, not only are prescriptions of the welfare school ruled out, but all prescription whatever. The economist as an adviser is completely stultified.” By the late 1940s, however, Robbins’ argument that economists should not deal with issues of inequality or distribution because they were normative and thus unscientific had catapulted itself to the heart of the economics profession. In so doing, economists began to present themselves as objective number crunchers whose only goal was to maximize productive efficiency in such a manner that reaches a Pareto optimal point, regardless of what the distributive ramifications may be. The question of distribution, they claimed, should be left to the political realm. In fact, students


were taught – through what came to be called the “second fundamental theorem of welfare economics” – that governments could, via lump-sum tax and transfers that did not distort the Pareto-optimizing wonders of the free market, determine what level of inequality they desired in their society. Crucially, however, such distributive discussions would be held by politicians – not economists – since it was a political and ethical issue rather than a scientific one. As a result, the second welfare theorem gave generations of neoclassical economists the perfect excuse to neglect the question of inequality.34

Once again, Samuelson’s textbook – which, it is important to remember, was clearly situated on the liberal side of neoclassical economics – offers a good articulation of this worldview by the 1960s:

“It is an ethical rather than a scientific question as to just how large, relatively, each person’s final income ought to be. As a science, economics can concern itself only with the best means of attaining given ends; it cannot prescribe the ends themselves. Indeed, if someone decided that he preferred a feudal-fascistic kind of society, in which all people with little black mustaches were to be given especially high incomes, the economist could set up the pricing rules for him to follow to achieve his strange design best.”35

The separation of economic inequality and economic efficiency was, perhaps, the most powerful force behind the marginalization of inequality economics in the twentieth century. It seeped into every nook and cranny of the discipline, while reaching the highest stages of economic decision-making. We have seen how, after Piketty’s book came out, Federal Reserve chief, Janet Yellen, warned of the dangers of inequality. In 2007, however, Fed chief and Princeton University economist Ben Bernanke made a very different argument, one that neoclassical economists had been making for more than half a century. In explaining why the Fed did not examine the issue of inequality, Bernake explained that he would “not draw any firm conclusions about the extent to which policy should attempt to offset inequality in economic outcomes; that determination inherently depends on values and social trade-offs and is thus properly left to the political process.”36

As Maurice Dobb and other historians of economic thought have shown, economists’ insistence to focus only on issues of efficiency would have a dramatic impact on the discipline. “Not only would egalitarian conclusions, distasteful to so many, be banished from sight,” Dobb explained, “but considerations about distribution… could apparently be banned from intrusion if not totally ignored.” Interestingly, however, most economists’ decision not to discuss distribution did not stop many of them from arguing – in what appears to be the very kind of normative claim that they derided – that free markets worked best for society. Thoughtful economists such as Paul Samuelson rarely spoke in such bombastic terms, but others were not so careful. And so, for instance, we can find in C.E. Ferguson’s 1969

35 Samuelson, Economics, 613.
textbook the argument that “a perfectly competitive, free enterprise system guarantees the attainment of maximum social welfare.”

Conclusion

Inequality economics today is witnessing a rebirth – but the depth and breadth of this intellectual impact on actual wealth and income inequality across the world remains to be seen. If this article on the past can tell us anything about the future, it is that for the economic study of inequality to really take off, and wealth inequities to significantly decline, the basic pillars of neoclassical economics may first have to be toppled.

Author contact: elicook@gmail.com

SUGGESTED CITATION:

You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-83/

On the current state of game theory

Bernard Guerrien  [l'Université Paris 1, France]

Copyright: Bernard Guerrien 2018
You may post comments on this paper at
https://rwer.wordpress.com/comments-on-rwer-issue-no-83/

“The formal language greatly limits the audience that really understands game theory; the abstraction blurs factors that natural thought takes into account and the formality creates an illusion that the theory is scientific” (Ariel Rubinstein, 2013).

Abstract

It emerges from a state-of-the-art review on game theory recently made by one of the most renowned specialists, Larry Samuelson, that the very idea of "applying" that theory to the study of concrete problems – in industrial economics for example – makes no sense. The restrictive nature of the underlying hypotheses is compounded by the problem of the multiplicity of "solutions" to its models. The countless "refinements" elaborated by game theorists to reduce the number of solutions only made matters worse while submerging game theory under the welter of mathematics – the arbitrary character of the choice of "the" solution is carried over the choice of "the" refinement. Arbitrariness also prevails in the so-called "evolutionary" approach of game theory, which has little to do with the initial idea of that theory (rational behaviors). The only area where game theory – in its "cooperative" form – had a certain utility was in the (prescriptive) field of matching between individuals who have an interest therein. The aim of the organizer-planner is not to find "the" optimal solution – which generally does not exist – but rather to do "for the best".

One of the most widespread myths in economics, but also in sociology and political science, is that game theory provides "tools" that can help solve concrete problems in these branches – especially in economics. Introductory and advanced textbooks thus often speak of the "applications" of game theory that are being made, giving the impression that they are revolutionizing the social sciences. But, looking more closely, we see that the few examples given concern mostly the usual "stories" (prisoners’ dilemma, "chiken", battle of sexes, entry deterrence, store chain paradox, centipede game, etc.) of "old" game theory. Take the four volume set Handbook of Game Theory with Economic Applications – a Handbook that provides an extensive account of what has been done in the field of game theory from its beginning, especially in economics, but not exclusively. Despite its title, there is not the slightest trace of a concrete example of an application, nor do we find any numerical data in its thousands of pages. This is not surprising. Mathematical reasoning requires clear and explicit enunciation of the assumptions used in its demonstrations. In particular, the assumptions concerning the information available to each player – his payoffs and those of the other players for each outcome of the game, the rules of the game, etc. – are so restrictive that there is no concrete situation in the world where they could possibly be verified, not even roughly (Guerrien, 2004, p. 2-3). As Ariel Rubinstein, another renowned game theorist puts it:

"Nearly every book on game theory begins with the sentence: ‘Game theory is relevant to …’ and is followed by an endless list of fields, such as nuclear strategy, financial markets, the world of butterflies and flowers, and intimate situations between men and women. Articles citing game theory as a source for resolving the world’s problems are frequently published in the daily press.
But after nearly forty years of engaging in this field, I have yet to find even a single application of game theory in my daily life” (Rubinstein, 2013, my italics).

Most serious game theorists know this but they are reluctant to admit it¹. In his recent overview of the subject “Game Theory in Economics and Beyond”, published in the Journal of Economic Perspective, Larry Samuelson, a prominent game theory specialist, gives a typical example of this contradictory [schizophrenic] attitude. Though it was probably not his intention, Larry Samuelson’s paper gives us a good picture of the impasse that game theory is in. After telling us that the theory, and its “tools” have become “essential”, – notably in economics where it leads to “extraordinarily useful models” –, the description that he gives of the state that the theory is in contradicts this rosy characterization. This is the case, for example, when he calls the mass of “refinements” of Nash equilibrium as “a menagerie”.

He is even less satisfied with the so-called “evolutionary” game theory, a branch that was “surrounded by a great deal of excitement” in the 1990s and has since “receded into the background.”

Apart from this contradictory attitude, Larry Samuelson, like many of the other prominent game theorists, makes some elementary logical mistakes. A typical example concerns his mischaracterization of Cournot’s model “solution” (equilibrium).

The Cournot’s “solution”: an inconsistent presentation

The only game model on which Larry Samuelson focuses in his paper is the old-fashioned Cournot duopoly (1838), typical of what he calls “classical” game theory. Even though he does not mention any concrete application of this model – not even a vague one as Cournot’s springs of mineral water –, L. Samuelson argues that it is “extraordinarily useful”. To describe the model, he draws two (reaction) curves and tells us that each player can deduce the equilibrium strategy (quantities of output) the same way “an analyst observing the game” can deduce it:

> “Under the classical view of game theory, one should be able to deduce the equilibrium play from the specification of the game and the hypothesis that it is commonly known that the players are rational. An analyst observing the game should be able to make such a deduction, as should the players in the game” (Samuelson 2016, p. 110, my italics).

This is a mistake since one of the main assumptions in Cournot's model is that each duopolist knows only his own cost function – from which he deduces his reaction function. Unlike “an analyst observing the game”, each player ignores everything about the other player. “Rational” players cannot therefore “deduce” the equilibrium strategy of the game, and equilibrium can only happen by accident (Bénicourt and Guerrien, 2008, p 320).

Larry Samuelson relies on his (erroneous) reasoning on Cournot’s model to justify the importance generally given to (Nash) equilibriums by game theorists:

¹ David Kreps also expresses skepticism, but less clearly than Ariel Rubinstein.
“This immediately answers an obvious question: Why are we interested in the equilibrium of a game? In the classical view, the equilibrium implication of a game will be obvious to rational players, and will just as obviously be reflected in their behavior” (Samuelson, 2016, p. 110, my italics).

He thereby falls under David Kreps’s criticism:

“When economic analysts invoke the notion of a Nash equilibrium, they are asserting, at least implicitly, that the situation in question has (or will have) a self-evident way to play. When, as often happens, they don’t say why it is that there is a self-evident way to play, then it is left to the reader either to supply the reason or to be suspicious of the results” (Kreps, 1991, p. 32).

The fact that the (Nash) equilibrium is not an obvious way to play (a way that is self-evident) is, for Kreps, the main weakness of game theory:

“The great weaknesses of game theory are that it is fuzzy (to say the least) on just when and, if so, why equilibrium analysis is relevant, and on what to do when equilibrium analysis is irrelevant” (Kreps, 1991, p. 36).

The “menagerie” of (Nash) equilibrium refinements

At the beginning of his article, Larry Samuelson misses out the problem of the relevance of Nash equilibrium as a “solution” (prediction) of game models. He prefers to focus on what he considers to be the major problem of game theory: the multiplicity of (Nash) equilibria.

Game theorists have addressed this issue by imposing additional restrictive conditions on equilibria – restrictions that they call “refinements” – but with little success. This is vividly described by Larry Samuelson:

“The equilibrium refinements literature was not a complete success. Instead of producing an equilibrium refinement that could command consensus, the literature gave rise to an ever-growing menagerie of refinements. New refinements tended to give rise to examples highlighting their weaknesses, followed by successor refinements destined to serve as the raw material for the next round. This seemingly endless cycle prompted Binmore (1992, p. 1) to liken the refinements quest to Hercules’ quest to kill the Hydra, with two new heads appearing in the place of each predecessor” (Samuelson, 2016, p. 111-112, my italics).

Carried away by his criticism, and in total contradiction with what he previously said about Cournot’s model, he questions the very relevance of Nash equilibrium:

“At the same time that many game theorists were busy inventing and reinventing refinements of Nash equilibrium, difficulties appeared in the attempt to show that at least Nash equilibrium, much less refinements of Nash equilibrium, could be deduced from the specification of the game and the hypotheses that the players are commonly known to be rational” (ibid, p. 112, my italics).
The quest for equilibrium refinements inevitably leads to the question of what exactly is a rational behavior, in particular in sequential games – such as the centipede game or the repeated “prisoner’s dilemma”. As Robert Aumann explains,

“the interesting aspect of the refinements is that they use irrationality to arrive at a strong form of rationality. In one way or another, all of them work by assuming that irrationality cannot be ruled out, that the players ascribe irrationality to each other with a small probability. True rationality (sic) requires ‘noise’; it cannot grow in sterile ground, it cannot feed on itself only” (Game Theory entry of The New Palgrave Dictionary of Economics, 2008).

Unfortunately, Aumann does not tell us anywhere what he means by “strong form of rationality” or by “true rationality” …

“Any behavior” can be “explained” in industrial economics

Larry Samuelson has equally harsh words for the use of game theory in industrial economics. “Heavily empirical” before the 1970s, industrial economics “became enthusiastically theoretical” in the 1980s, thereby succumbing to the trend of game theory. For Larry Samuelson, the systematic recourse to (noncooperative) game theory, with its countless (Nash) equilibrium refinements, can explain anything:

“Strategic models came to be used to explain price discrimination, advertising, entry deterrence, limit pricing, and a host of other phenomena. The difficulty was that an impression soon formed that a sufficiently determined modeler could construct a model explaining any behavior, no matter how counterintuitive” (Samuelson, 2016, p. 118).

It is perhaps for this reason that:

“the strategic revolution in industrial organization did not maintain its momentum” (ibid, p. 118).

After giving credit to Thomas Schelling for pointing out the importance of “the personal and cultural context”, L. Samuelson mentions Keynes and emphasizes the need for “careful work on the art of choosing models”. However, he does not provide any successful example of this “art” apart from a rapid reference to the hypothetical problem between a president of the United States and a “Middle Eastern dictator” whose rationality is not ensured…

And about Jean Tirole?

Jean Tirole, we are told, won the Nobel Prize in economics because:

“he has made important theoretical research contributions in a number of areas, but most of all he has clarified how to understand and regulate industries with a few powerful firms” (Press release of the Economic Sciences Prize Committee of the Royal Swedish Academy of Sciences, October 2014).
He was supposedly also rewarded for his proposing a “unified theory” based on “his analysis of firms with market power” (ibid).

Although the issue of firms with “market power” is at the very origin of “industrial economics”, Larry Samuelson’s article – which was published two years after the award of the Nobel Prize to Tirole – says nothing about this “unified theory”, in which he apparently does not believe. Neither does he mentions any paper published by Jean Tirole. He only blames him for failing to speak of the cooperative approach in his game theory textbook (Fudenberg and Tirole, 1991).

Firms with market power – whether on the selling side or on the buying side – will meet at some time and they will need to bargain. What does game theory have to say about this?

**And the bargaining problem, “basic point of departure for studying economics”?**

After reminding us that:

> “Edgeworth identified the bargaining problem as the basic point of departure for studying economics”,

Larry Samuelson observes that, despite the efforts made by game theorists,

> “economic reasoning was not helpful in identifying which of the typically many efficient outcomes might appear” (Samuelson, 2016, p. 121).

Therefore,

> “one cannot easily point to examples where bargaining methods have been overhauled in response to game theoretic insights” (ibid, p. 121).

The difficulty stems from the fact that the various procedures proposed by game theorists for overcoming the indeterminateness that characterizes bargaining situations are “too sensitive to fine details of the model”.

Among these “details”, there are

> “the timing of offers and counteroffers, the specification of the information structure, the length of the horizon, the length of a time period, and other details” (ibid, p. 121).

Each of these details can have a significant impact on the “outcome” of the model, regardless of the solution concept used.

For Larry Samuelson:

> “given a choice from a collection of models that give sharply different results, with little guidance as to which is appropriate, it is not surprising that one might avoid using such models” (ibid, p. 122).
Therefore, there is not much to expect from game theory when it comes to the bargaining problem.\(^2\)

A similar statement can be made about evolutionary game theory which enjoyed its days of glory at the end of the 20\(^{th}\) century.

**Evolutionary game theory: another disappointment**

Samuelson reminds us that during the 1990s, evolutionary game theory was “surrounded by a great deal of excitement” in the economics profession (p 115). There was hope that it could remove the indeterminateness related to the multiplicity of Nash equilibria.

Instead of deducing “the equilibrium actions from the structure of the game”, players can proceed “to a trial-and-error process that (one hopes) tends to push them toward equilibrium” (Samuelson, 2016, p. 115).

This seems reasonable at first sight... but it is not. Rational individuals notice that their forecasts are mistaken and progressively modify them by taking into account what they observe during the process. Hence, the equilibrium that they reach – if they ever reach it... – depends on what happened all along the path that they followed to reach it (it is “path dependent”) – so it cannot be deduced from the specifications of the initial model.

Samuelson illustrates the idea of an evolutionary game by returning to Cournot’s model:

“Evolutionary game theory puts the dynamic process back into the picture. Interestingly, Cournot (1838) motivated his equilibrium for the duopoly as the limiting outcome of a best-response-based adjustment process” (ibid, p 115).

This justification of equilibrium only makes sense if we assume that firms behave like robots that are programmed to make offers and counteroffers on the basis of a reaction function given once and for all – whereas “conscious” players modify their behavior in accordance with what they observe during the process.\(^3\)

The evolutionary approach radically alters the essence of game theory. It does so, not by introducing “dynamics”,\(^4\) but by turning players into robots whose “algorithm” is given by a list of instructions – that game theorists call a strategy. Players do not choose strategies, each of them is (reduced to) a strategy. A game is therefore given by a set of strategies and rules that specify the order and the number of times (“rounds”) those strategies meet in pairwise contests.

---

\(^2\) Ebay’s case is notable. Built originally around the idea that “appropriate prices” (whatever that may mean) would emerge from generalized bargaining, the website had to change its policy: nearly 90% of the exchanges are now taking place with a price set by the sellers. There are no absolute rules that prevent (some) players’ “strategic behavior” from destroying confidence in the game.

\(^3\) Cournot’s reaction function is based on the conjecture that “the other player will not modify his offer if I modify mine” – while it is what he does all along the equilibrium “search” process.

\(^4\) In the introduction to their *Theory of Games and Economic Behaviour*, Von Neumann and Morgenstern explain that it will take “at least a century” to introduce dynamics into game theory.
In principle, the outcome of an evolutionary game is simple to determine: feed the (given) strategies (lists of instructions) and rules of the game into a computer and look at the gains of each player (strategy) – gains can be expressed in terms of offsprings or in anything else.

Even if they radically differ in their conception, the “conscious choice” and the “robotic behavior” approaches of game theory are both unfortunately often thrown together into the same pot. This confusion is maintained because the concept of Nash equilibrium and that of evolutionary stable strategy (which is specific to the evolutionary approach) share a certain formal similarity. In both cases, any “deviation” from “equilibrium” by a player – a modification in his choice or a (random) “mutation” in his (unique) strategy – results in a lower payoff for him.

Samuelson eventually notes that after the “excitement” of the 1990s – as attested by the countless “tournaments” organized with various alternatives of the win-win game – evolutionary game theory has progressively “receded into the background” (Samuelson, 2016, p. 116).

This can be interpreted as a “success”. However, in retrospect, it appears that it is a meager success. The evolutionary approach has, in particular, failed to solve the problem of the multiplicity of Nash equilibria that cripples game theory:

“Evolutionary models do not consistently lead to any of the standard refinements of Nash equilibrium, much less produce a consensus on what a useful refinement might be” (ibid, p. 116).

Larry Samuelson gives an example – perhaps an opportunity to mention one of his publications… – in which the “evolutionary dynamics” do not eliminate the (weakly) dominated strategies whereas they “should be eliminated in every (Nash) equilibrium refinement”. This would be sufficient to explain why evolutionary theory “has receded into the background” (ibid, p. 116).

Some successes? Auctions and mechanism design

For Larry Samuelson,

“Two of the obvious successes of game theory are auctions and matching” (ibid, p. 121).

In both cases, the theoretician adopts the point of view of someone who seeks to organize the relationships between the participants in the best way – on a voluntary basis or in return for a compensation. Therefore, the aim is not to explain how things are but rather how they should be, according to a criterion specific to each situation.

After mentioning the “obvious successes” of auction theory, Larry Samuelson acknowledges that:

“modeling an auction gives rise to a seemingly endless series of choices— are values common or idiosyncratic, are the bidders risk neutral or risk averse, is there a resale market, will the bidders collude, are the bidders
symmetric, and so on—again without definitive indications as to which is the obvious modeling choice” (ibid, p. 122).

Once again, we end up “without definitive indications as to which is the obvious modeling choice” (ibid, p 122).

The game theorist can propose the type of auction that he deems appropriate to a specific situation. However, he must make a certain number of assumptions – notably concerning the behavior of the bidders. There can be no guarantee that these assumptions are verified and therefore that the solution proposed is the “best” one – whatever that may mean. The person who effectively organizes the auction can, as Samuelson says, only gain “useful insights” from it. The “obvious success” in the application of game theory to auctions is not so obvious…

Auctions are a specific case of “mechanism design”, an approach that proceeds in the “reverse order” of what is usually done in standard game theory: the game theorist sets a goal – a function to be maximized for example – and then seeks (“designs”) the rules of the game (“the mechanism”) that would allow the goal to be achieved, or at least to get as close to it as possible. In the case of an auction organized by a seller, the goal can be to obtain the highest price – the one that the bidder with the highest values is willing to pay. Mechanism design is therefore closer to the approach adopted by an engineer – or a central planner that establishes rules and incentive systems – than to that of a person who seeks to explain what happens or to predict what will happen.5

Matching is, according to L. Samuelson, the other “obvious success” of game theory. The organizing “engineer” now decides how the players are “matched” by trying to do what is best according solely to their preferences. His problem is, at first glance, relatively simple because he does not need to care – as in the case of an auction – about the players’ behavior, the information available, their beliefs, etc. The matching problem is quite common: students that seek to get into the “best schools” and schools that wish to recruit the “best students”, marriage candidates, organ donors and recipients… The aim of the organizer is to match people in such a way that there is no match (A, B) by which both A and B would be individually better off than they are with the element to which they are currently matched.6

This stability property – no one has an incentive to replace his partner in the absence of (another) person that is inclined to do so – is used to characterize certain concepts of solution within game theory. Simple in principle, the determination of stable matchings is complicated in practice because matchings are derived from rankings – and not from numbers that have a meaning in themselves. This “technical” difficulty is compounded by an “ethical” issue: the organizer must generally choose between several stable matchings and then “favor” certain players at the expense of others. The theorist or the practitioner can simply find the “solutions” (stable matchings), give them to the organizer and let him choose the one he prefers.

5 The model of perfect competition – with its “center” that “gives” prices and changes them until their equilibrium value is found, etc. – can be considered as an example of mechanism design intended to achieve a form of social welfare (Pareto optimality). For this aim to be achieved, agents must act as price takers and there must be a complete system of markets.

6 Matching is often associated with the idea of a “market” – for marriage, for organ donation, etc. A very strange market, indeed, without prices and where the participants merely inform the “organizer” about their possibilities or preferences…
Does research into matching techniques fall in the field of game theory? Not if one considers that game theory deals with “decision-makers whose actions mutually affect one another” (Aumann, 2008) since there is only one “decision-maker” in the case of matching – the “engineer” or the organizer – who seeks to solve a problem related to what we commonly call “operational research”. Accordingly, there is no trace of the “matching” issue found in game theory textbooks nor in the several popular texts related to this field.

Is the “cooperative” approach of game theory a road to salvation?

Research into “satisfying” matchings falls within the so-called “cooperative” – “coalitional” to be more precise – game theory. Larry Samuelson blames game theorists for abandoning the cooperative approach in the 1970s, in favor of the non-cooperative one. The cooperative approach actually held a prominent place in Von Neumann and Morgenstern’s seminal book *Theory of Games and Economic Behavior* (1944). Andrew Schotter, a Morgenstern student, explains that

> “von Neumann and Morgenstern (VNM) were looking for a way to break the circularity of the ‘I think he thinks that I think’ logic of strategically interdependent situations. They wanted to provide a way for players to behave that was independent of their expectation of what their opponent intended to do” ([Schotter](#), 1992, p. 106).7

Hence the special attention VNM gave what they called “orders of society” – *institutional arrangements* in current language. Multiplicity of orders simply reflects the variety of existing societies, where individuals form groups (coalitions) in which they “cooperate”. As is suggested by the word “order”, theoreticians are particularly interested in situations where groups maintain themselves – where they are stable.8 The other cooperative solution concepts – VNM stable sets, bargaining sets, kernel, core, nucleolus, etc... – are in the same vein.

Larry Samuelson does not ask why the non-cooperative approach of game theory – despite its impasses – almost entirely replaced the cooperative approach in the 1970s. The most plausible explanation is ideological, as often in economics. Indeed, the non-cooperative approach – which implies that each solution must be derived from individual maximizations – is consistent with the canons of “microeconomic foundations”, the sole scientific or rigorous approach accepted by the consensus established in the 1970s within the mainstream paradigm. Unhappily, the only great “success” that can be claimed by game theory in practice – matching, which is a type of planning – is not really in conformity with this new zeitgeist …

**Conclusion**

Initially developed by mathematicians interested in situations such as those of certain parlor games, with specific rules, game theory progressively became a label for all situations, real or imaginary, in which the gains of each participant, in money or in anything else, depends, at

---

7 The “minimax” strategy for zero-sum games, the only non-cooperative concept in *Theory of Games and Economic Behavior*, supposes that “a certain return is guaranteed, no matter what one’s opponent does” (p. 106, Schotter’s italics). It aims at safety rather than maximum gains.

8 Schotter believes that, given his “Austrian” convictions, Morgenstern would have been reluctant to use game theory for “social planning” and for “institutional engineering”, as mechanism design does.
least partly, on the choices made by the other participants. Since most activities in social life correspond to this type of situation, the scope of game theory is, in principle, unlimited. In practice, the constraints that the use of mathematics requires implies that this scope is reduced to more or less complex “fables” or “proverbs” where the requirements concerning the information available to the players are so unreal that they are never verified in practice, not even roughly.

As Ariel Rubinstein clearly says in Frankfurter Allgemeine:

“game theory is a collection of fables and proverbs. Implementing a model from game theory is just as likely as implementing a fable. A good fable enables us to see a situation in life from a new angle and perhaps influence our action or judgment one day. But it would be absurd to say that “The Emperor’s New Clothes” predicts the path of Berlusconi …” (Rubinstein, 2013).

Bibliography


Author contact: bguerrien@sfr.fr

SUGGESTED CITATION:

You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-83/
Why game theory never will be anything but a footnote in the history of social science
Lars Pålsson Syll [Malmö University, Sweden]

"I have no fellow-feeling with those economic theorists who, off the record at seminars and conferences, admit that they are only playing a game with other theorists. If their models are not intended seriously … why do they expect me to spend my time listening to their expositions? Count me out of the game" (Robert Sugden).

Game theory is an axiomatic-mathematical theory that presents a set of axioms that people have to ‘satisfy’ by definition to count as ‘rational.’ This makes for ‘rigorous’ and ‘precise’ conclusions – but never about the real world. Game theory does not give us any information at all about the real world. Instead of confronting the theory with real-world phenomena it becomes a simple matter of definition if real-world phenomena are to count as signs of ‘rationality.’ It gives us absolutely irrefutable knowledge – but only since the knowledge is purely definitional.

"Mathematical theorems are tautologies. They cannot be false because they do not say anything substantive. They merely spell out the implications of how things have been defined. The basic propositions of game theory have precisely the same character" (Binmore, 1994: 23).

Pure game theorists, like Ken Binmore, give us analytical truths – truths by definition. That is great – from a mathematical and formal logical point of view. From a scientific point of view, however, it is rather uninteresting and uninformative. Even if game theory gives us ‘logical’ truths, that is not what we are looking for as scientists. We are interested in finding truths that give us new information and knowledge of the world in which we live.

Scientific theories are theories that ‘refer’ to the real-world, where axioms and definitions do not take us very far. To be of interest for an economist or social scientist that wants to understand, explain, or predict real-world phenomena, ‘pure’ theories have to be ‘interpreted’ – they have to be ‘applied’ theories. A game theory that does not go beyond proving theorems and conditional ‘if-then’ statements – and do not make assertions and put forward hypotheses about real-world individuals and institutions – is of little consequence for anyone wanting to use theories for real-world purposes.

Although the critique put forward in this essay will be predominantly of a methodological ilk, much of it will also be substantive in nature. And much of what is discussed does not only apply to game theory, but to a large extent also to ‘rational choice’ theory. The reason is simple. Game theory rests on ‘rational choice’ theory. A deep, fundamental, critique of game theory has to be directed against the very foundations of its assumptions. Uncritically taking for granted and accepting the axiomatic status of those assumptions brings the critique to a halt, or forces the critique to address problems and anomalies of a second-order magnitude. The most reasonable procedure is arguably to follow Robert Solow’s suggestion (Klamer ed. 1984: 146):
“Suppose someone sits down where you are sitting right now and announces to me that he is Napoleon Bonaparte. The last thing I want to do with him is to get involved in a technical discussion of cavalry tactics at the battle of Austerlitz. If I do that, I’m getting tacitly drawn into the game that he is Napoleon. Now […] like nothing better than to get drawn into technical discussions, because then you have tacitly gone along with their fundamental assumptions; your attention is attracted away from the basic weakness of the whole story. Since I find that fundamental framework ludicrous, I respond by treating it as ludicrous – that is, by laughing at it – so as not to fall into the trap of taking it seriously and passing on to matters of technique.”

Game theory's rational fools

Game theory is, like mainstream economics, model-oriented. There are many reasons for this – the history of the discipline, having ideals coming from the natural sciences (especially physics), the search for universality (explaining as much as possible with as little as possible), rigour, precision, etc. Most mainstream economists and game theorists want to explain social phenomena, structures and patterns, based on the assumption that the agents are acting in an optimizing (rational) way to satisfy given, stable and well-defined goals.

The procedure is analytical. The whole is broken down into its constituent parts so as to be able to explain (reduce) the aggregate (macro) as the result of the interaction of its parts (micro). Building their economic models, modern mainstream economists ground their models on a set of core assumptions describing the agents as ‘rational’ actors and a set of auxiliary assumptions. Together these assumptions make up the base model of all mainstream economic models. Based on these two sets of assumptions, they try to explain and predict both individual and social phenomena.

The core assumptions (cf. Pålsson Syll, 2016b) typically consist of completeness, transitivity, non-satiation, expected utility maximization, and consistent efficiency equilibria.

When describing the actors as rational in these models, the concept of rationality used is instrumental rationality – choosing consistently the preferred alternative, which is judged to have the best consequences for the actor given his in the model exogenously given interests and goals. How these preferences, interests, and goals are formed is not considered to be within the realm of rationality, and a fortiori not constituting part of economics proper.

The picture given by this set of core assumptions – ‘rational choice’ – is a rational agent with strong cognitive capacity that knows what alternatives she is facing, evaluates them carefully, calculates the consequences and chooses the one – given his preferences – that she believes has the best consequences according to her. Weighing the different alternatives against each other, the actor makes a consistent optimizing choice and acts accordingly.

Besides the core assumptions the model also typically has a set of auxiliary assumptions that spatio-temporally specify the kind of social interaction between ‘rational’ actors that take place in the model. These assumptions can be seen as giving answers to questions such as: who are the actors and where and when do they act, which specific goals do they have, what are their interests, what kind of expectations do they have, what are their feasible actions, what
kind of agreements (contracts) can they enter into, how much and what kind of information do they possess, and how do the actions of the different individuals interact with each other.

So, the base model basically consists of a general specification of what (axiomatically) constitutes optimizing rational agents and a more specific description of the kind of situations in which these rational actors act (making the auxiliary assumptions serve as a kind of restriction of the intended domain of application for the core assumptions and the deductively derived theorems). The list of assumptions can never be complete since there will always be unspecified background assumptions and some (often) silent omissions (usually based on some negligibility and applicability considerations). The hope, however, is that the ‘thin’ list of assumptions shall be sufficient to explain and predict ‘thick’ phenomena in the real, complex, world.

These models are not primarily constructed – especially not in game theory – for being able to analyze individuals and their aspirations, motivations, interests, etc., but typically for analyzing social phenomena as a kind of equilibrium that emerges through the interaction between individuals.

Now, of course, no one takes the base model (and the models that build on it) as a good (or, even less, true) representation of reality (which would demand a high degree of appropriate conformity with the essential characteristics of the real phenomena, that, even when weighing in pragmatic aspects such as ‘purpose’ and ‘adequacy,’ it is hard to see that this ‘thin’ model could deliver). The model is typically seen as a kind of thought experimental ‘as if’ benchmark device for enabling a rigorous mathematically tractable illustration of social interaction in an ideal-type model world, and to be able to compare that ‘ideal’ with reality. The ‘interpreted’ model is supposed to supply analytical and explanatory power, enabling us to detect and understand mechanisms and tendencies in what happens around us in real economies.

Based on the model – and on interpreting it as something more than a deductive-axiomatic system – predictions and explanations can be made and confronted with empirical data and what we think we know. The base model and its more or less tightly knit axiomatic core assumptions are used to set up further ‘as if’ models from which consistent and precise inferences are made. If the axiomatic premises are true, the conclusions necessarily follow. But if the models are to be relevant, we also have to argue that their precision and rigour still holds when they are applied to real-world situations. They often do not. When addressing real economies, the idealizations and abstractions necessary for the deductivist machinery to work simply do not hold.

If the real world is fuzzy, vague and indeterminate, then why should our models build upon a desire to describe it as precise and predictable? The logic of idealization, that permeates the base model, is a marvellous tool in mathematics and axiomatic-deductivist systems, but a poor guide for action in real-world systems, where concepts and entities are without clear boundaries and continually interact and overlap.

Being told that the model is rigorous and amenable to ‘successive approximations’ to reality is of little avail, especially when the law-like (nomological) core assumptions are highly questionable and extremely difficult to test. Being able to construct ‘thought-experiments’ depicting logical possibilities does not take us very far. An obvious problem with the mainstream base model is that it is formulated in such a way that it realiter is extremely difficult to empirically test and decisively ‘corroborate’ or ‘falsify.’ Such models are from a
scientific-explanatory point of view unsatisfying. The ‘thinness’ is bought at too high a price, unless you decide to leave the intended area of application unspecified or immunize your model by interpreting it as nothing more than two sets of assumptions making up a contentless theoretical system with no connection whatsoever to reality.

Ontology – the lacking dimension

What is lacking in the overly simplistic mainstream view on using mathematical modelling in game theory is an ontological reflection on the conditions that have to be fulfilled for appropriately applying the methods of mathematical modelling. Using formal mathematical modelling, one sure can guarantee that the conclusion holds given the assumptions. However, there is no warrant that the validity we get in abstract model worlds automatically transfer to real-world economies. Validity and consistency may be good, but it is not enough.

From a realist perspective, both relevance and soundness are *sine qua non*. In their search for validity, rigour and precision, game theorists construct models that standardly assume things like ‘perfect information,’ ‘consistently aligned beliefs,’ backward induction,’ ‘common knowledge,’ etc., etc. At the same time, the models standardly ignore things like complexity, diversity, genuine uncertainty, and expectations formation. Behavioural and experimental economics – not to speak of psychology – show beyond any doubts that peoples’ preferences, choices, and forecasts, are regularly influenced by factors that are not incorporated into game theory.

So, in what way can one maintain that these models give workable foundations for explaining social interaction between different actors and decision makers? In game theoretical models – where the conclusions follow deductively – mathematics is the preferred means to assure that we get what we want to establish with deductive rigour and precision. The problem, however, is that what guarantees this deductivity are as a rule the same things that make the external validity of the models wanting. The core assumptions are not very many, and so, if the modellers want to establish ‘interesting’ facts about the economy, they have to make sure the set of auxiliary assumptions is large enough to enable the derivations. But then – how do we validate that large set of assumptions that give the game theorist her ‘clarity’ and ‘consistency’ outside the model itself? How do we evaluate those assumptions that are used for no other purpose than to guarantee an analytical-formalistic use of mathematics? And how do we know that our model results ‘travel’ to the real world?

On a deep level, one could argue that the one-eyed focus on internal validity and consistency make game theory irrelevant since its insistence on deductive-axiomatic foundations does not earnestly consider the fact that its formal logical reasoning, inferences and arguments show an amazingly weak relationship to their everyday real-world equivalents. Although the formal logic focus may deepen our insights into the notion of internal validity, the rigour and precision have a devastatingly important trade-off: the higher the level of rigour and precision, the smaller is the range of real-world application. The more game theoreticians insist on formal logical validity, the less they have to say about the real world.

Back in 1991, when earning his first PhD with a dissertation on decision making and rationality in social choice and game theory, yours truly concluded (Syll 1991:105) that

“repeatedly it seems as though mathematical tractability and elegance – rather than realism and relevance – have been the most applied guidelines
for the behavioural assumptions being made. On a political and social level, it is doubtful if the methodological individualism, ahistoricity and formalism they are advocating are especially valid.”

Mainstream colleagues were – to say the least – not exactly überjoyed. But twenty years later, renowned game theorist Ariel Rubinstein (2012b), confirmed the doubts about the value of game theory:

“Game theory is about a collection of fables. Are fables useful or not? In some sense, you can say that they are useful, because good fables can give you some new insight into the world and allow you to think about a situation differently. But fables are not useful in the sense of giving you advice about what to do tomorrow, or how to reach an agreement between the West and Iran. The same is true about game theory.”

Expected utility theory and the behavioural challenge

In game theory, preferences are standardly expressed in the form of an expected utility function. Although the expected utility theory has been known for a long time to be both theoretically and descriptively inadequate, game theorists gladly continue to use it, as though its deficiencies were unknown or unheard of. But when models are plainly wrong, you have better replace them. As Matthew Rabin and Richard Thaler (2001: 230) put it:

“It is time for economists to recognize that expected utility is an ex hypothesis, so that we can concentrate our energies on the important task of developing better descriptive models of choice under uncertainty.”

In a similar vein, Daniel Kahneman (2011) and Richard Thaler (2016) maintain that expected utility theory is seriously flawed since it does not take into consideration, e.g., the basic fact that people’s choices are influenced by changes in their wealth. Where standard game theory assumes that preferences are stable over time, behavioural economists have forcefully again and again shown that preferences are not fixed, but vary with different reference points. How can a theory that does not allow for people having different reference points from which they consider their options have a (typically unquestioned) axiomatic status within economic theory?

Much of what experimental and behavioural economics come up with, is really bad news for mainstream economic theory, and to just conclude, as many mainstream economists do, that game theoretical insights can be applied to most decision-making areas, sounds, to say the least, somewhat lame, when the works of people like Rabin, Thaler and Kahneman, show that expected utility theory is nothing but transmogrifying truth.

If we cannot show that the mechanisms or causes we isolate and handle in our models are stable, in the sense that what when we export them from our models to our target systems they do not change from one situation to another, then they only hold under ceteris paribus conditions and a fortiori are of limited value for our understanding, explanation and prediction of our real-world target system.
Ken Binmore and other game theorists try to ‘save’ game theory by treating it as an axiomatic system and making all its claims into tautologies – ‘true’ by the meaning of propositional connectives. The problem is, of course, that ‘saving’ theories and models by this kind of immunizing strategy are totally unacceptable from a scientific point of view. If game theory has nothing to say about the real world, why should we care about it? As long as no convincing justification is put forward for how the inferential bridging between model and reality _de facto_ is made, game theoretical model building is little more than hand-waving. The real challenge is to acknowledge and face real-world uncertainty and still try to explain why economic transactions and social interaction take place – instead of simply conjuring the problem away by assuming things like ‘common knowledge’ and ‘perfect information,’ or treating uncertainty as if possible to reduce to stochastic risk, or by immunizing models by treating them as purely deductive-axiomatic systems.

**Nash equilibrium**

Nash equilibrium has since it was introduced back in the 1950’s (cf. Nash 1951) come to be the standard solution concept used by game theorists. The justification for its use has been mainly built on dubious and contentious assumptions like ‘common knowledge’ and individuals exclusively identified as instrumentally rational. And as if that was not enough, one actually, to ‘save’ the Holy Equilibrium Grail, has had to further make the ridiculously unreal assumption that those individuals have ‘consistently aligned beliefs’ — effectively treating different individuals as incarnations of the microfoundationalist ‘representative agent.’

“According to the way we normally use the common knowledge assumption along with that of symmetrically rational, and, for that matter, perfectly rational individuals, each and every individual is assumed to reason the same way about the game. We in effect have reduced the problem of reasoning in an interactive situation to the reasoning of a representative ideal individual who knows the game in full and shares this knowledge by virtue of the common knowledge assumption with each and every other participant. The game theorist and the participants in the game are in the same situation. Everybody comes exactly to the same conclusions as everybody else when thinking about the game before the specific play of the game starts.

In sum, as far as the reasoning itself is concerned we are not talking about some interactive reasoning practice. It is rather an ideal type of reasoning to which all ideal type reasoners are assumed to ‘converge.’ It is the reasoning of a representative ideally rational individual” (Hartmut Kliemt, 2009:145f).

In the beginning — in the 1950s and 1960s — hopes were high that game theory would enhance our possibilities of explaining the behaviour of interacting actors in non-parametric settings. And this is where we ended up! A sad story, indeed, showing the limits of methodological individualism and instrumental rationality.

So why not give up on the Nash concept altogether? Why not give up the vain dream of trying to understand social interaction by reducing it to something that can be analyzed with models of instrumentally interacting ideally rational individuals?
“We believe that a variety of contributory factors can be identified … It is possible that the strange philosophical moorings of neoclassical economics and game theory have played a part. They are strange in at least two respects. The first is a kind of amnesia or lobotomy which the discipline seems to have suffered regarding most things philosophical during the postwar period … The second is the utilitarian historical roots of modern economics … Thirdly, the sociology of the discipline may provide further clues … All academics have fought their corner in battles over resources and they always use the special qualities of their discipline as ammunition in one way or another. Thus one might explain in functionalist terms the mystifying attachments of economics and game theory to Nash” (Varoufakis & Hargreaves-Heap 1995: 108).

When criticising game theory you often get the rather uninformative and vacuous answer that we all have to remember that game theory – as is mainstream economics at large – is nothing but an ‘as if’ theory built on ‘as if’ rationality. But as Ariel Rubinstein (2012a:53) has it, this however only shows that “the phrase ‘as if’ is a way to avoid taking responsibility for the strong assumptions upon which economic models are founded.”

The mathematical-deductivist straitjacket used in game theory presupposes atomistic closed-systems – i.e., something that we find very little of in the real world, a world significantly at odds with an (implicitly) assumed logic world where deductive entailment rules the roost. Ultimately then, the failings of game theory have their roots in a deficient ontology. The kind of formal-analytical and axiomatic-deductive mathematical modelling that makes up the core of mainstream economics is hard to make compatible with a real-world ontology. A game theory that is relevant to the world in which we live can never achieve the same degree of rigour and precision as in logic, mathematics or the natural sciences. In game theory, with its addiction to the deductivist approach of formal-mathematical modelling, model consistency trumps real-world coherence. That certainly is getting the priorities wrong. Creating models for their own sake is not an acceptable scientific aspiration – impressive-looking formal-deductive (mathematical) models should never be mistaken for truth.

On the limited applicability of game theory

Many mainstream economists – still – think that game theory is useful and can be applied to real-life and give important and interesting results (cf., e.g., Hausman 2005). That, however, is a rather unsubstantiated view. What game theory does is, strictly seen, nothing more than investigating the logic of behaviour among non-existant robot-imitations of humans. Knowing how those ‘rational fools’ play games do not help us to decide and act when interacting with real people. Knowing some game theory may actually make us behave in a way that hurts both ourselves and others (cf. Frank et al. 1993). Decision-making and social interaction are always embedded in socio-cultural contexts. Not taking account of that, game theory will remain an analytical cul-de-sac that never will be able to come up with useful and relevant explanations.

“Imagine you and someone you do not know can share $100. It is up to you to propose how to divide the $100 between the two of you, and the other player will need to accept or reject your proposal. If he rejects the proposal, neither of you will receive anything. What sum will you offer the other player?
I have data on the choices of about 12,300 people, most of them students, who were asked this question. Nearly half of the participants (49%) offered the other player the fair offer of $50 …

The participants in the experiment who make the embarrassing offer of just $1 because they learned this in a game theory course are again the distinguished members of the Victims of Game Theory organization. And if they played the game in real life, their achievements would be inferior to those who had not become wise by studying game theory” (Rubinstein 2012a:111f).

Over-emphasizing the reach of instrumental rationality and abstracting away from the influence of many known to be important factors, reduces the analysis to a pure thought experiment without any substantial connection to reality. Limiting theoretical economic analysis in this way – not incorporating both motivational and institutional factors when trying to explain human behaviour – makes economics insensitive to social facts.

“For certain specific, local problems, game theory is a very nice way of thinking about how people might try to solve them, but as soon as you are dealing with a general problem like an economy or a market, I think it is difficult to believe that there is full strategic interaction going on. It is just asking too much of people. Game theory imposes a huge amount of abstract reasoning on the part of people … That is why I think game theory, as an approach to large scale interaction, is probably not the right way to go” (Kirman, 2011: 53).

Game theorists extensively exploit ‘rational choice’ assumptions in their explanations. That is probably also the reason why, as argued by Guala (2006:239), game theory has not been able to “accommodate the anomalies in its theoretical framework.” That should hardly come as a surprise to anyone. Game theory with its axiomatic view on individuals’ tastes, beliefs, and preferences, cannot accommodate very much of real-life behaviour. It is hard to find really compelling arguments in favour of us continuing down its barren paths since individuals obviously do not comply with, or are guided by, game theory. Apart from (perhaps) few notable exceptions – like Schelling’s (1978) and Akerlof’s (1970) explanations (although, as argued in (Rosenberg 1995:ch 6), actually only suggesting what might be the rationale behind these phenomena) of segregation and ‘lemons’ – it is difficult to find really successful applications of game theory. Why? To a large extent simply because the boundary conditions of game theoretical models are false and baseless from a real-world perspective. And, perhaps even more importantly, since they are not even close to being good approximations of real-life, game theory is lacking predictive power. This should come as no surprise. As long as game theory sticks to its ‘rational choice’ foundations, there is not much to be hoped for.

In an interview, Ariel Rubinstein (2012b) had the following to say on the question of the real-world value of game theory:

“Is game theory useful in a concrete sense or not? Game theory is an area of economics that has enjoyed fantastic public relations. [John] Von Neumann [one of the founders of game theory] was not only a genius in mathematics, he was also a genius in public relations. The choice of the name “theory of games” was brilliant as a marketing device …
I think it’s a very tempting idea for people, that they can take something simple and apply it to situations that are very complicated, like the economic crisis or nuclear deterrence. But this is an illusion … I believe that game theory is very interesting. I’ve spent a lot of my life thinking about it, but I don’t respect the claims that it has direct applications.

The analogy I sometimes give is from logic. Logic is a very interesting field in philosophy, or in mathematics. But I don’t think anybody has the illusion that logic helps people to be better performers in life …

In general, I would say there were too many claims made by game theoreticians about its relevance. Every book of game theory starts with “Game theory is very relevant to everything that you can imagine, and probably many things that you can’t imagine.” In my opinion that’s just a marketing device …

I have not seen, in all my life, a single example where a game theorist could give advice, based on the theory, which was more useful than that of the layman …”

Game theorists can, of course, marginally modify their tool-box and fiddle with the auxiliary assumptions to get whatever outcome they want. But as long as the ‘rational choice’ core assumptions are left intact, it seems a pointless effort of hampering with an already excessive deductive-axiomatic formalism. If you do believe in a real-world relevance of game theoretical ‘science fiction’ assumptions such as expected utility, ‘common knowledge,’ ‘backward induction,’ correct and consistent beliefs etc., etc., then adding things like ‘framing,’ ‘cognitive bias,’ and different kinds of heuristics, do not ‘solve’ any problem. If we want to construct a theory that can provide us with explanations of individual cognition, decisions, and social interaction, we have to look for something else.

In real life, people – acting in a world where the assumption of an unchanging future does not hold – do not always know what kind of plays they are playing. And if they do, they often do not take it for given, but rather try to change it in different ways. And the way they play – the strategies they choose to follow – depends not only on the expected utilities, but on what specifics these utilities are calculated. What these specifics are – food, water, luxury cars, money etc. – influence to what extent we let justice, fairness, equality, influence our choices (cf. Yaari & Bar-Hillel 1984). ‘Welfarism’ – the consequentialist view that all that really matters to people is the utility of the outcomes – is a highly questionable short-coming built into game theory, and certainly detracts from its usefulness in explaining real-life choices made outside the model world of game theory.

Games people play in societies are usually not like games of chess. In the confined context of parlour-games – like in the nowadays so often appealed to, for ‘defending’ the usefulness of game theory, auction negotiations – the rather thin rationality concept on which game theory is founded may be adequate. But far from being congratulatory, this ought to warn us of the really bleak applicability of game theory. It is hard to see how the chess playing experience would help us in any substantial way to understand and explain strategic interaction between individuals in real-world social contexts. Game theory, with its highly questionable assumptions on ‘rationality’, equilibrium solutions, information, and knowledge, simply makes it useless as an instrument for explaining real-world phenomena.
Applications of game theory have on the whole resulted in massive predictive failures. People simply do not act according to the theory. They do not know or possess the assumed probabilities, utilities, beliefs or information to calculate the different (‘subgame,’ ‘trembling-hand perfect’) Nash equilibria. They may be reasonable and make use of their given cognitive faculties as well as they can (cf. Pålsson Syll 2007:ch 7), but they are obviously not those perfect and costless hyper-rational expected utility maximizing calculators game theory posits. And fortunately so. Being ‘reasonable’ make them avoid all those made-up ‘rationality’ traps that game theory would have put them in if they had tried to act as consistent players in a game theoretical sense.

The lack of successful empirical application of game theory shows there certainly are definitive limits of how far instrumental rationality can take us in trying to explain and understand individual behaviour in social contexts. The kind of preferences, knowledge, information and beliefs – and lack of contextual ‘thickness’ – that are assumed to be at hand in the axiomatic game theoretical set-up do not give much space for delivering real and relevant insights of the kind of decision-making and action we encounter in our everyday lives.

**Where did game theory go wrong?**

Instead of making formal logical argumentation based on deductive-axiomatic models the message, we are arguably better served by social scientists who more than anything else try to contribute to solving real problems – and in that endeavour, other inference schemes may be much more relevant than formal logic.

“The weaknesses of social-scientific normativism are obvious. The basic assumptions refer to idealized action under pure maxims; no empirically substantive law-like hypotheses can be derived from them. Either it is a question of analytic statements recast in deductive form or the conditions under which the hypotheses derived could be definitively falsified are excluded under ceteris paribus stipulations. Despite their reference to reality, the laws stated by pure economics have little, if any, information content. To the extent that theories of rational choice lay claim to empirical-analytic knowledge, they are open to the charge of Platonism (Modellplatonismus). Hans Albert has summarized these arguments: The central point is the confusion of logical presuppositions with empirical conditions. The maxims of action introduced are treated not as verifiable hypotheses but as assumptions about actions by economic subjects that are in principle possible. The theorist limits himself to formal deductions of implications in the unfounded expectation that he will nevertheless arrive at propositions with empirical content. Albert’s critique is directed primarily against tautological procedures and the immunizing role of qualifying or ‘alibi’ formulas. This critique of normative-analytic methods argues that general theories of rational action are achieved at too great a cost when they sacrifice empirically verifiable and descriptively meaningful information” (Habermas, 1988:48).

Game theoretical models build on a theory that is abstract, unrealistic and presenting mostly non-testable hypotheses. One important rationale behind this kind of model building is the quest for rigour, and more precisely, logical rigour. Instead of basically trying to establish a
connection between empirical data and assumptions, ‘truth’ has come to be reduced to a question of fulfilling internal consistency demands between conclusion and premises, instead of showing a ‘congruence’ between model assumptions and reality. This has, of course, severely restricted the applicability of game theory and its models.

The world in which we live is inherently uncertain and quantifiable probabilities are the exception rather than the rule. To every statement about it is attached a ‘weight of argument’ that makes it impossible to reduce our beliefs and expectations to a one-dimensional stochastic probability distribution. If “God does not play dice” as Einstein maintained, I would add “nor do people.” The world as we know it has limited scope for certainty and perfect knowledge. Its intrinsic and almost unlimited complexity and the interrelatedness of its organic parts prevent the possibility of treating it as constituted by ‘legal atoms’ with discretely distinct, separable and stable causal relations. Our knowledge accordingly has to be of a rather fallible kind.

To search for precision and rigour in such a world is self-defeating, at least if precision and rigour are supposed to assure external validity. The only way to defend such an endeavour is to take a blind eye to ontology and restrict oneself to prove things in closed model-worlds. Why we should care about these and not ask questions of relevance is hard to see. We have to at least justify our disregard for the gap between the nature of the real world and the theories and models of it.

If the real world is fuzzy, vague and indeterminate, then why should our models build upon a desire to describe it as precise and predictable? Even if there always has to be a trade-off between theory-internal validity and external validity, we have to ask ourselves if our models are relevant.

‘Human logic’ has to supplant the classical, formal, logic of deductivism if we want to have anything of interest to say of the real world we inhabit. Logic is a marvellous tool in mathematics and axiomatic-deductivist systems, but a poor guide for action in real-world systems, in which concepts and entities are without clear boundaries and continually interact and overlap. In the world in which we live, we are better served with a methodology that takes into account that usually the more we know, the more we know we do not know.

**Taking uncertainty seriously**

Game theory has created its own ‘as if’ parallel world. To judge the whole game theoretical project one has to evaluate the relation between that model world and the real world. To our understanding of the way the world works, belongs a thorough recognition of the ontological restriction of the kind of cognitional and epistemological assumptions we can make about what kind of knowledge and information individuals can possibly have. It is important, not least since in a game theoretical – strategic – setting, knowledge and information are essential parts of the context in which individuals make their decisions and act.

Like the ‘rational choice’ theory on which it is based, game theory postulates that individuals are rational in the specific definitional meaning that they decide and act on given preferences through calculating expected utilities. To be able to make those calculations it is assumed that – besides actors having very strong logical and cognitional capacities – uncertainty conditions can be treated as risk equivalents.
Here you can, of course, question all the ingredients in that ‘rationality’ definition. But, for the moment, let us focus on the uncertainty issue, since if the reductionist view on uncertainty is untenable in real-world contexts, it is also not possible to consider individuals as behaving according to the definitional rationality. Facing genuine uncertainty, actors cannot, strictly seen, make any rational calculations at all, but rather have to (partly) base their decisions and acts on institutions, rule following, imitation, norms, conventions, ‘animal spirits’, etc., etc.

Facing genuine uncertainty the player in a game, when trying to act optimally given his preferences, cannot be sure that the genuinely uncertain – \textit{ex ante} – consequences he decides and acts on will actually – \textit{ex post} – materialize. In repeated games, game theorists usually assume that the new information that the players accumulate during the game will somehow solve the uncertainty problem. But it does not, at least not in a non-ergodic world, where the arrow of time makes the future – and the materialized consequences – different to the past.

“If the actions that I undertake in t0 will have very different consequences according to the eventual state of the world in t1, it is crucial to gather reliable knowledge about these states. But how could I evaluate in t0 my beliefs about the state of the world in t1? If the world were repetitive (governed by immutable laws) and these laws were known, I could assume that what I find out about the present state is relevant to determine how the future state (the one that will prevail) will be. It would make then sense to apply a strategy for gathering empirical evidence (a sequence of actions to collect new data). But if the world is not repetitive, what makes me think that the new information may be at all useful regarding future events? ...

Conceiving economic processes like sequences of events in which uncertainty reigns, where consequently there are “no laws”, nor “invariants” or “mechanisms” to discover, the kind of learning that experiments or past experience provide is of no use for the future, because it eliminates innovation and creativity and does not take into account the arboreal character and the open-ended nature of the economic process ... However, as said before, we can gather precise information, restricted in space and time (data). But, what is the purpose of obtaining this sort of information if uncertainty about future events prevails?” (Marqués 2016: 118-9).

The effects of taking the concept of genuine uncertainty seriously are indeed far-reaching. Living in a world permeated by unmeasurable uncertainty – not quantifiable stochastic risk – forces us to make decisions based on anything but ‘rational expectations’ and ‘expected utility.’ In a genuinely uncertain world, we have to base our expectations and calculations on the confidence or ‘weight’ we put on different events and alternatives. The expectations we form and the calculations we make are (partly) based on weighing probabilities by ‘degrees of belief,’ beliefs that often have precious little to do with the kind of stochastic probabilistic calculations made by the rational agents as modelled by game theory. Often we ‘simply do not know,’ and to assume – as is standard in ‘rational choice’ and game theory – that what has worked before, will continue to do so in the future, is unwarranted. One cannot simply project history onto the future.

Robert Lucas (1981: 223-4) once wrote that “in cases of uncertainty, economic reasoning will be of no value.” Now, if that was true, it would put us in a tough dilemma. If we have
to consider – as Lucas – uncertainty incompatible with economics being a science, and we actually know for sure that there are several and deeply important situations in real-world contexts where we – both epistemologically and ontologically – face genuine uncertainty, well, then we actually would have to choose between reality and science. And that cannot be right. We all know we do not know very much about the future. We all know the future harbours lots of unknown unknowns. Those are ontological facts we just have to accept. Looking the other way and assume a lot of known to be utterly and ridiculously unreal things – as in game theory – is not the right way to tackle the problems uncertainty poses for social sciences.

Under uncertainty, individuals make mistakes, and not only of the usual ‘white noise’ kind assumed by ‘rational choice’ and game theory, but even systematic ones, because of the inherent uncertainty of the probability judgments individuals have to base their decisions and acts on in real-world contexts without the possibility to reduce uncertainty to risk. In real-world situations – where there are few, if any, ‘nomological machines’ – every day is different and without much hopes of making improvements by learning. Of course, you may learn things, in a historical sense, about things that have already taken place, but since there is no possibility of learning much about the future in an uncertain non-ergodic world without stability and invariance, the learning you do will not help you make decisions and act on ‘rational’ calculations. In an uncertain world ontologically characterized by non-repetitiveness and emergence, there is very little that can justify the assumptions of expected utility calculations on which ‘rational choice’ and game theory is founded. With uncertainty, the definitional rationality that game theory presumes simply is not to be had. And hence game theory becomes irrelevant. Irrelevant and useless.

Game theory – not really explaining decision-making and acting at all

In game theory, although it ‘describes’ strategic interaction between ‘rational’ individuals, the decisions made by those individuals are always seen as completely independent decisions. Game theory is not really interested in if individuals actually behave in accordance with its axioms and assumptions. It only cares about the possibility of describing the behaviour as if they maximize their expected utility.

Game theoretical rationality is at its core defined as a player being able to choose the ‘best’ action given the beliefs she has about the beliefs, preferences and possible actions of other players. The individual depicted in game theory is an unbearable inhuman robot-imitation, but in real life we all err, and we do it in systematic ways (cf. Kahneman, 2011; and Thaler, 2016).

So – again – why assume individuals are ‘rational”? Mostly because game theorists want to be able to make behaviour predictable. It would have been much better if they had stuck more to real-life persons and accepted that these to a large extent are unpredictable creatures trying to cope with living in a largely genuinely uncertain world. Game theory takes for granted that every player has the ability to put herself in the shoes of other players. In reality, that is not so. One size does not fit all, and that goes for game-theoretical shoes too.

There do exist innumerable considerations that influence decision makers and that are contrary to the ‘rational choice’ assumption used in game theory. And with the restricted conception of human behaviour and far-reaching tractability assumptions on which game theory builds, most people are probably forgiven for not regarding its achievement as a big deal. Its ‘rational choice’ foundations – with its definitional consistency-reduced view of
‘rationality’ – makes much of it a platitude. Telling us that individuals choose the preferred alternatives in their preference orderings is rather uninformative – and empirically shown over and over again to be far from true. The close link that is presumed between expected utility and the choices individuals make is often severed in real-world settings. No man is an island. Actors and decision makers – as repeated experiments with ‘dictator’ and ‘ultimatum’ games have shown (cf. Carmichael (2005), Rubinstein (2012), Frank et al. (1993), Hausman (2005), Kahneman et al. (1986) – often do care about commitments, which involve them in making counter-preferential non-expected-utility-maximizing choices based on ethics, solidarity, mutual trust, duty, obligation, morals, norms, social positions, reputation, rules of conduct, imagination, gut feeling, ambiguity, etc., etc. Individuals do – more or less often – choose to act in ways that give them less personal expected utility than other available alternatives. What an individual regards as good from a ‘social’ point of view may not be the same as what he regards as good from a ‘private’ point of view. Individuals are often ambiguous and succumb to akrasia. Individuals do frequently make unselfish choices. Although our choices and acts may have bad anticipated consequences, we sometimes prefer to go beyond the limited and mute ‘rationality’ that consequentialist expected utility calculations prescribe. Real-world individuals have richer and more complex preferences than those that are posited in the game theoretical behavioural axioms.

“The traditional theory has too little structure. A person is given one preference ordering, and as and when the need arises this is supposed to reflect his interests, represent his welfare, summarize his idea of what should be done, and describe his actual choices and behavior. Can one preference ordering do all these things? A person thus described may be ‘rational’ in the limited sense of revealing no inconsistencies in his choice behavior, but if he has no use for these distinctions between quite different concepts, he must be a bit of a fool. The purely economic man is indeed close to being a social moron. Economic theory has been much preoccupied with this rational fool decked in the glory of his one all-purpose preference ordering. To make room for the different concepts related to his behavior we need a more elaborate structure” (Sen, 1977: 335f).

As we have seen, there simply are no ‘objective’ probability distributions out there for individuals to build expectations and utility calculations on. That also makes the kind of calculations and predictions that game theory presupposes more or less impossible. Keynes’ (1936) famous beauty test illustrates part of the problem. To solve this kind of almost self-referential vicious circle reasoning, instrumental logic has to be supplanted by other kinds of logic if we want to be able to explain the final expectations on which people decide to act on.

Game theoretical ‘solutions’ critically presuppose that players are rational and know that other players are rational, and so on. The moment the slightest doubt about that knowledge about other people’s beliefs (and, strictly seen, the process through which they are reached) creeps in, the theory could become seriously misleading.

Most of the mathematical reasoning (especially on the cognitive capacities of the players) that lies behind game theoretical results are beyond the formal comprehension of most real-world decision makers and players. If ‘rational’ at all, most of them are only ‘boundedly’ so and frequently use different kinds of pragmatic ‘heuristics’ or ‘rules of thumb.’
Beliefs and preferences *per se* do not in any substantial meaning explain anything, unless one is able to show that those beliefs and preferences also are the *de facto* driving causal reasons behind the decisions and acts of individuals. That also applies to game theory. Finding people in game-like situations acting like game theory predicts does not *per se* constitute a vindication of game theoretical explanations. The players may have acted out of quite different reasoning and beliefs than those stipulated by game theory. If so, game theory still does not explain the behaviour in any meaningful way.

The ideally ‘rational’ agent in game theory is an extremely idealized one. In our imperfect and genuinely uncertain world we never – never – run into anyone equipped with that kind of consistent preferences, information, knowledge or calculating abilities. We are humans and not ‘players’ in the ideal-type world of game theory. If we are found to maximize any expected utilities at all, it is (probably) only by chance.

In game theorists’ famous apopthegm, words are nothing but ‘cheap talk.’ Motivation is purely forward-looking and bygones are bygones. But in real-life situations, we do usually find people having a duty or obligation to stand by their words. History matters and investments we have made in the past do influence our present and future behaviour. Reasons can be backwards-looking. Promises made have a binding power on most of us. That game theory, footed on its instrumental rationality concept, cannot accommodate those facts, only shows that its notion of rationality is too restricted.

‘Common knowledge’ means that anything known to anyone is also known by everyone else. From a realist point of view, this is such a ridiculous assumption to make, that one wonders how anything concluded from this assumption could ever be imagined to travel from the model world to the real world. It is an assumption dictated purely out of tractability considerations, and without that assumption, game theory cannot come up with equilibrium solutions to many of its games. But so what? More than anything else it just underlines how useless the theory is for explaining real-world phenomena.

Since game theory does not incorporate real-world psychological or social factors, what *realiter* rule or influence many decisions and acts of individuals are *de facto* treated as totally irrelevant. In that sense, and contrary to what many economists applying game theory may think, game theory is not at all a theory of how individuals make decisions and act. It is not a descriptive theory. It is rather a normative deductive theory telling us that if we want to behave ‘rational’ we have to do so in a way consistent with game theory. But – people seldom act in the way prescribed by game theory. They do not cohere with all or any of those ‘rationality’ assumptions, and so the real normative force of game theory is close to zero. Even if this to game theorists is of no interest – the ‘if’ clause is not satisfied, and so the ‘then’ clause may be whatever – negative empirical observations certainly detract from the purported value of game theory. If the impressive precision of the game theoretical solutions has to be bought at such a high price, most people are probably not prepared to pay for being considered ‘irrational’.

**Game theoretical obscurantism**

In game theory, agents do not only have beliefs about each other, but also have to have beliefs about beliefs, and so on. To somehow short-circuit the infinite regress problem inherent in this strategic uncertainty situation, one standardly appeals to the equilibrium
notion. That is, as we already know, a standard tractability assumption in mainstream economics for closing the models, but even this evasion does not work when we have multiple equilibria (cf. Hargreaves-Heap ed. 1992: 107f).

Individuals outside these ‘as if’ models – with their purely instrumentalist justification – simply do not possess the cognitive capacity to make the kind of expected utility calculations game theory presupposes. Although this is a *sine qua non* for constructing the games analyzed in game theory, it has very little justification or warrant for analyzing real-world behaviour or decision making. The game theoretical ‘as if’ model results simply do not bridge to the real world.

So why should we care about ‘as if’ results derived in the extremely narrow framework of a game theory erected on ‘rational choice’ pillars? Why, indeed, should we, in the words of Elster (2015:453), care about mending an ‘obscurantist’ theory with its

“uncanny combination of mathematical sophistication on the one hand and conceptual naïveté and empirical sloppiness on the other, [and in which] the mathematics, which could have been a tool, is little more than toy?”

You could probably try to give answers in terms of the aesthetics of mathematical model-building, but it would be more fair to just admit that, from the point of view of social science, there is absolutely no reason at all why we should care about game theory simply because it is of no value at all – beyond some few very restricted contexts such as, e.g., auctions – for explaining real-world decision making and behaviour. Science fiction can sure function as an inspiration to all of us, but to use it as a building block for a relevant and realistic social science is hardly tenable. “Mit der Dummheit kämpfen Götter selbst vergebens,” but we still have to leave the pointless excessive deductive-axiomatic formalism of game theory behind, and instead reorient our endeavours into building a more modest, relevant, realistic and robust social science.

Outside the confines of the model world, game theoretical findings seem to have very little relevance. Under genuine uncertainty, game theory does not offer any advice at all, since game theory is founded on assumptions that are known to be patently surreal in most interesting economic contexts. No wonder then that its real-world value has to be seriously questioned – models that only make sense when we accept assumptions that over and over again have been shown by psychologists and behavioural scientists to be ridiculous, is no more valuable to us than Walt Disney fictions.

All theories and models have to use sign vehicles to convey some kind of content that may be used for saying something of the target system. But purpose-built tractability assumptions – like the modelling assumption that “an analysis of complex structures by parts is possible” (Kliemt, 2009: 125) – made solely to secure a way of reaching deductively validated results in mathematical models, are of little value if they cannot be validated outside of the model.

Models do not only face theory. They also have to confront the world. But being able to model a ‘credible world,’ a world that somehow could be considered real or similar to the real world, is not the same as investigating the real world. Questions of external validity, the claims the extrapolation inference is supposed to deliver, are important. It can never be enough that models somehow are regarded as internally consistent. One always also has to pose
questions of consistency with the data. Internal consistency without external validity is worth nothing.

The links between real-life situations and the abstract formulation of problems in game theory are often difficult to discern. The values of individualism and competitiveness that are embedded in game theory are in many decision contexts at odds with reality. If individuals’ decisions and beliefs cannot be reduced to a common measuring stick – expected utility – they cannot be translated into numbers, and so ‘disappear’ from the game theoretical analysis (cf. Rapoport 1962:113). Ignoring much of what really influence individual behaviour in real-life situations, applying the narrow and artificial perspective of game theory outside the model world is more likely to obscure the deliberations behind individuals’ decisions and acts than to reveal them. Game theory simply does not elucidate real-life problems. On the contrary. It is probably mostly unhelpful and harmful. Analyzing and trying to explain individual interaction in complex social contexts is a daunting task. Solving it by making absurd ‘as if’ simplifying assumptions, as in game theory, cannot, however, be the right solution. By applying game theory in real-world contexts we often end up thinking like game theorists and incorporate the values inherent in the theory – with often terrible results (as shown e.g. by the experiments reported in Frank et al., 1993).

In many situations what matters is not outcomes, but intentions. Fairness considerations do influence how we decide and act in many situations, and if we think those we ‘play’ with are pursuing unfair strategies, we are as a rule prepared to let them pay for that, even if it also ‘hurts’ ourselves. Even if that behaviour does not comply with game theoretical ‘rationality,’ it certainly complies with most people’s ideas about fairness and reciprocity.

Game theory may devise models – ‘nomological machines’ – in which it is possible to derive law-like regularities: ‘Satisfying all the core and auxiliary assumptions in the base model, players will decide …. act … and do …’ The problem with these logical model deductions is, of course, that the game theoretical assumptions and results do not in any obvious way represent or relate to real-life situations. The game theoretical results may be law-like, rigorous, precise, and exact – but what good does that do if it comes at the cost of real-world irrelevance? If the assumptions on which game theory builds do not fit, to any considerable degree, with the world around us, well, then we certainly have to wonder what use is game theoretical ‘Glasperlenspiel.’

Conclusion

Heavy use of formalism and mathematics easily foster the view that a theory is scientific. But although game theory may produce ‘absolute truths’ in imaginary model worlds, in the real-world the game theoretic models are nothing but – as Rubinstein (2012a) puts it – fables. Fables much reminiscent of the models used in logic, but also like them, delivering very little of value for social sciences trying to explain and understand real-life phenomena. The games that game theory portrays are model constructs, models without significant predictive capacity simply because they do not describe an always much more complex and uncertain reality.

Being at its heart a sub-discipline within pure mathematics, game theorists are not overly concerned with whether game theory represents real-world phenomena. Fine. But since most social scientists are of a different opinion, game theorists also have to accept that to most
social scientists, game theory is deemed useless for explaining interaction between individuals in the real world.

According to Morgenstern (1964:8) game theory was “designed to give meaning to what common sense vaguely calls rational behavior.” It is difficult to concur. Game theory may be very rigorous, but it certainly also has many evident shortcomings and defects. As the famous ‘prisoner’s dilemma’ poignantly shows, game theory has deep problems explaining social facts by its individualistic and egocentric ‘rational choice’ models (cf. Luce & Raiffa, 1957; Rapoport & Chammah, 1965; Kreps, 1990; Axelrod, 1984). Social interaction cannot be exhaustively described as strategic interaction. Posited behavioural regularities are nothing but illegitimate generalizations based on taking for granted that a theory that is able to explain has to be a universal theory with an unlimited domain as long as the core assumptions of ‘rationality’ are fulfilled. Without other forms of interaction, society as we know it would not be possible. Maybe this is the only real usefulness, if any, of game theory: it shows the severe limits of ‘rational choice’ and strategic interaction in explaining social interaction.

Although some economists consider it useful to apply game theory and use game theoretical definitions, axioms, and theorems and (try to) test if real-world phenomena ‘satisfy’ the axioms and the inferences made from them, we have argued that that view is without warrant. When confronted with the real world we can (hopefully) judge if game theory really tells us if things are as postulated. The final court of appeal for models is the real world, and as long as no convincing justification is put forward for how the inferential bridging de facto is made, model building is little more than hand-waving that give us rather little warrant for making inductive inferences from the model world to the real world.

The real challenge in social science is to accept uncertainty and still try to explain why different kinds of transactions and social interactions take place. Simply conjuring problems away by assuming patently unreal things and treating uncertainty as if it was possible to reduce to stochastic risk, is like playing tennis with the net down. That is not the kind of game that scientists working on constructing a relevant and realist science want to play.

Half a century ago there were widespread hopes game theory would provide a unified theory of social science. Today it has become obvious those hopes did not materialize. This ought to come as no surprise. Reductionist and atomistic models of social interaction – such as the ones mainstream economics and game theory are founded on – will never deliver sustainable building blocks for a realist and relevant social science. That is also the reason why game theory never will be anything but a footnote in the history of social science.

References


Morgenstern, Oskar (1964). On some criticisms of game theory


Pålsson Syll, Lars (2016b). When the model becomes the message – a critique of Rodrik. *real-world economics review*.


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

SUGGESTED CITATION:

You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-83/
The creation of jobs
Michael Joffe [Imperial College, London, UK]

Abstract
This paper investigates the source of jobs in the modern economy, excluding the state, non-profit and financial sectors. The approach is centered on the firm and its profitability, and in particular, proposes a real investment led economy perspective. Rather than assuming that productive systems already exist, as is frequently done, it examines how they are established and renewed, and also how they may cease to exist. Investment is seen as the strategic decision process that establishes the number and type of jobs for the medium term, and their approximate wage level. Thus, at the time of the job creation decision, it is not clear who the potential workers are – and particularly when technological change and off-shoring are involved, even what population they will be drawn from. Investment often involves a business plan: a joint decision about the product, location, technology, scale of production, etc, as well as employment. In addition, tactical decisions are taken during the course of production. Both types of decision take account of the economic environment, and involve Knightian uncertainty. Unemployment exists when fewer jobs are created than there are people who would like to fill them, because there are too few investment opportunities that are perceived as being potentially profitable. In the investment decision, the whole package needs to be coherent, and offering lower wages may not be enough to make an investment idea potentially profitable. This explains why non-frictional unemployment can occur, and persist, especially when perceived investment opportunities are few. It involves asymmetry: a shortage of workers is reflected in an increase in the wage level, whereas a shortage of jobs is manifest in terms of unemployment – quantity not price. The real investment led economy perspective is also able to account for employment changes in six major types of scenario: a new company, major technological change, relocation, plant closure, regional decline and a major depression. Existing theories struggle to explain these phenomena. In addition, this perspective naturally addresses several puzzles in standard labor economics.

JEL codes J21, J23, J31, J60, E24, L21

Key words firm-centered approach, real investment led economy, unemployment, technological change, offshoring, labor market puzzles, wage flexibility puzzle, unemployment volatility puzzle, employer size-wage effect (ESWE)

1. A long term dynamic view of the labor market

A fundamental issue in economics is the source of jobs in a modern capitalist economy: where do they come from? Who creates them? When? Why? How? A comprehensive theory of the labor market needs to be able to answer these questions in a satisfactory manner.

In addition to the context of business as usual, such a theory must engage with the major situations that affect employment, unemployment and wage levels. As well as accurately describing routine operation, it must be able to give a good account of the implications for employment and wages under conditions that involve alteration of the trajectory of a firm or an economy. This means the ability to encompass the following scenarios, four of which represent the key strategic decisions that firms take, the other two being broader conditions that impact firms and their profitability:

1 I would like to thank Victoria Chick and Jon Fjeld for helpful discussions and comments on a previous draft.
This paper introduces a new perspective on employment, unemployment and wage setting that can naturally meet this challenge. It is centered on the firm and its decision making, rather than on the relationship between firms and workers, because it is the firm that takes the major initiatives that affect the location, quantity, type, quality and remuneration of jobs. The central consideration for a firm is its profitability, as survival is impossible for long when making a loss. The primary focus here is on the investment process undertaken by firms from time to time, including the setting up of new firms. This viewpoint could be called the real investment led economy perspective. It is derived from a description of the observed behavior of firms, especially the time order of key decisions and events, and is evidence-based in this sense. The term “real” is included to specify investment in the real economy, rather than financial or real estate investments – the focus is on productive firms. The description applies not only within manufacturing, but also to private sector non-financial services. I am neglecting the public sector and non-profit firms in this paper.

The major traditions in labor economics, based e.g. on the neoclassical theory of the firm and on the search-and-matching approach, are well established. But they share the feature that a production system is assumed to be already in existence. In the case of neoclassical theory, the focus is on the decision to employ an extra worker, or one fewer, at the margin, given a production set-up that is already a going concern. In the search-and-matching perspective, current employment positions and prospective workers are brought together, which requires the assumption that the jobs already exist, and the workers are appropriate for them. This implies a limitation in both cases, which is particularly clear in a situation where a firm may decide between different options for locating the new production, or where there are substantially different possible technologies that have major implications for the number and type of workers that would be required. Under those conditions, it is not even obvious who the prospective workers are, until the location and/or technology decision is taken.

The assumption that employment and wages can be validly analyzed in the context of a productive setup that already exists could be justified for a relatively simple economy, for example medieval and early modern England, e.g. where one is discussing the hiring of different types of building workers (Clark, 2005). However, in the modern economy most types of occupation have been introduced by capitalist firms at some stage, often combined with new technology and/or new products. This was true of employees in the early 19th century cotton mills of Lancashire, of Henry Ford’s production line workers, and remains true today, e.g. of workers in Shenzhen assembling iPhones. The modern production system continually creates new forms of employment, as well as new technology and new products. Even for the large majority of capitalist firms that are not innovative in this respect, and are merely treading the already-established paths, their investments are the primary source of the private sector jobs created in the real economy. Accordingly, the real investment led economy view traces the source of employment opportunities to the initiatives taken by capitalist firms, even when they are relatively routine ones.
The real investment led economy perspective maintains that investment is endogenous to the system, an integral part of it. Although investment may appear to be an autonomous decision, it responds to external stimuli in the economic environment – the institutional and technological environment, and especially the projected demand for the product. This latter provides a link with aggregate demand at the macro level. A hallmark of this viewpoint is that it is dynamic: it is centrally concerned with how economies change over time. The analysis focuses as much on how productive systems come into existence, and how they change, as on more short-term questions about how they operate.

The real investment led economy perspective is the positive side of the more general firm-centered view. For relatively strong firms in a favorable situation, the focus is on investment. The negative side applies to firms in a weak position, resulting from their lack of capacity and/or the lack of demand for their product(s). It similarly emphasizes the central causal importance of the firm and its decisions, the primary aim being to make profits or at least to avoid losses.

2. An outline of the real investment led economy perspective

The core idea in the real investment led economy perspective is that the real economy has been created over time by the initiatives taken by firms. This occurs in the process of investment, either by existing firms or by start-up companies. Investment is the occasion on which those who control the firm make the key strategic decisions that set the pattern for its future operation. This is a single many-faceted decision, nowadays often specified in a business plan. At the time that the plan is developed, it is based on the best information that can reasonably be obtained, and can therefore be assumed to be optimal in this sense. The centrality of investment is a characteristic of the modern “capitalist” economy.

The decisions include the nature of the product(s), location, technology, broad scale of production, and types of labor – the numbers required of each skill type and an expectation of their approximate wage levels. One of the primary concerns is recruitment and retention of suitable workers, in the right numbers, at the right wage, in the proposed location. A business plan also analyses the availability of finance, both internal and external, and estimates the potential sales quantity and price. The wage and price decisions are made in comparison with what others are doing, including the going wage for each type of employment. Ideally, it also takes into account how this is likely to evolve.

Crucially, all these components need to fit together in a way that promises success – the sums need to add up, and the strategy needs to be coherent. In particular, decisions involving employment type and quality cannot be separated from those on technology. For example, there is evidence that investment in information technology involves also the adjustment of work practices as well as the rethinking of product offerings, a process that requires experimentation (Bresnahan et al., 2002). And various components of workers’ skill are related to firms’ technological inputs (Abowd et al., 2007).

The concept of investment is broader than the standard image of purchasing equipment for the purpose of manufacturing. It includes starting a hotel or restaurant, or setting up a homecare agency, or an airline.
Not all investment is aimed at increasing (or transforming) productive capacity. Auxiliary investment, for example a marketing initiative aimed at increasing sales of a current product line, enhances the value of an existing investment in productive capacity. This can also affect employment.

An investment therefore involves far more than just buying some machines. This contrasts with the thought experiment frequently found in textbooks, in which a firm decides separately on a fixed and a variable factor of production, usually capital and labor, respectively.

All aspects of the investment are set in accordance with the requirement of future profitability. This will clearly depend on expected future market conditions, relating both to the inputs to production and to the product, that can only be estimated rather roughly at the time that the investment is being planned due to the presence of Knightian uncertainty. This central idea is based on the non-controversial statement that firms invest so as to maximize expected future profit, taking into account the economic environment.\(^2\)

The economic environment includes the types and quality of workers available in particular candidate locations, and the likely cost of employing the required number for the proposed scale of production. The major employment decision therefore rests on the potential productivity of the available workers compared with the wage level required to attract them. The going wage must be taken as given, taking account of the possibilities of offering extra in order to enhance worker quality, motivation and retention, and of augmenting existing human capital by training programs – seen in the broader context that includes location and technology. In this way, firms create employment positions, or in the terminology of Acemoglu and Autor (2011), occupations which are bundles of tasks.

3. The real investment process

Firms invest when the expectation of future gains exceeds the expectation of the necessary costs by an amount that generates a satisfactory return on capital. Costs here include fixed costs as well as variable costs, and taxes. Investment is a strategic decision that sets the firm’s future direction. It includes the following:

1. Employment depends on the firm’s investment decision – this neglects the cases in which a new position is created in response to an initiative taken by a potential worker, which may sometimes occur (especially in the case of very highly-skilled/specialized people), but is probably unusual.

2. Numbers of workers of each skill type are similarly part of the firm’s investment decision. This is a joint decision with the technology decision, e.g. how many machines are to be purchased.

3. The approximate wage level for each type of worker is also part of the firm’s investment decision, taking account of the going wage, plus a possible increment to improve recruitment, motivation and retention if that can be afforded. This may involve a joint

---

\(^2\) Actually, maximization is a convenient assumption but is not an essential feature; this perspective could be compatible with e.g. a satisficing view of firms’ behavior.
decision with location, especially for large firms that are able to (re)locate abroad, to a country with a beneficial ratio of worker productivity to wage level.

4. Investment depends on the availability of finance, including retained profits as well as possible external funding.

The wage level is often specified as a wage range, i.e. it is approximate ex ante, giving scope for negotiation with workers once they are identified, and/or with their representatives, which corresponds to the wage-bargaining situation in the literature; on the other hand, when a precise wage is specified as a condition of employment, this corresponds to wage posting (Hall and Krueger 2008). In this perspective, they are not regarded as polar opposites – rather, the difference is the comparatively minor distinction between approximate and precise specification of the wage level, respectively. Firms also make the decision on how to structure the employment relationship in such a way as to try and best motivate employees (Oyer and Schaefer, 2011).

The ability to make a potentially profitable investment depends on the firm’s capacities, including its managerial capacity (Penrose, 1959), and its strength relative to others in the context of the power struggle between competing firms (Joffe, 2011). With relatively weak firms, the situation can arise that a firm may go for years without investing, i.e. there is stasis and possible decline. This could be for lack of funds and/or possible investment opportunities that promise improvement, or it could just be a question of inadequate or complacent management. The consequence for workers is likely to be stagnation in wages, and ultimately insecurity of employment if the firm then struggles to survive. Similarly, in a sector that is disrupted by the introduction of substantially lower costs or new/improved products, incumbent firms need to react, e.g. by trying to adopt the new practices, or by scaling back production and/or restructuring. These conditions affect firms that are in a weak position relative to competitors – they are on the back foot – and are therefore less able to take initiatives that are potentially profitable. Whilst strictly speaking not within the real investment led economy viewpoint, it is still best seen in a firm-centered perspective, with a primary focus on the degree of profitability.

4. Decision making during the course of production

In addition to the strategic decision making involved in investment, more day-to-day tactical decisions are subsequently taken in the course of production. These include recruiting workers to replace those who have decided to leave; dealing with disciplinary issues and conflict; maintenance and repair of equipment and premises; and responding to current market conditions, e.g. by reducing the price of items that are not selling.

The production-related component includes the following items that involve the labor market:

1. The exact wage level may be subject to negotiation with workers once contact is made with them. Generally speaking, workers with higher-level or scarce skills tend more often to be able to negotiate, because they are in a better bargaining position than easily-replaceable low-skilled workers (Hall and Krueger, 2008).

3 The topic of the relationship between competitor firms, and in particular their relative strength, is beyond the scope of the current paper – for a discussion, see Joffe [2011], especially section 3.1.
2. The wage level may be subject to negotiation in the context of new matches during the life course of an investment, i.e. workers recruited at some later date, for example to replace departed employees; this is likely to be related to market conditions such as the unemployment rate (Pissarides, 2009). Renegotiation with existing workers may also occur, e.g. in an annual review.

3. The number of hours worked per worker is similarly subject to negotiation/renegotiation and to current market conditions.

Decisions that are not directly related to the labor market similarly have both strategic and tactical components, e.g. the quantity and price of the output, and the sourcing of raw materials.

Any satisfactory account of the process of job creation needs to combine the strategic decisions with the tactical modifications that occur during the lifetime of the investment. These latter typically involve workers (individually and/or via their representatives) as well as firms, and are amenable to a more conventional approach, e.g. using the canonical search and matching framework (Pissarides, 2000).

From the firm’s viewpoint, production decisions are taken in the light of the economic environment, just as investment ones are. For example, a firm may decide to increase wages in order to poach or retain workers, in relation to the actions of other firms (Moscarini and Postel-Vinay, 2008). The “economic environment” here includes the financial position of the firm – what capital it can raise in addition to its own retained funds – and market conditions, i.e. what it will be able to sell, as well as supply-chain factors, and what competitors are doing.

Both investment and production decisions involve the passage of time: firms make their decisions on the basis of their expectations of what is likely to be profitable, and workers then respond. They also both involve Knightian uncertainty. However, they differ greatly both in their time horizon and in their degree of uncertainty.

Tactical decisions in the course of production tend to have a time horizon of days, weeks or months, and the degree of uncertainty in the outcome is moderate. Strategic investment decisions are accompanied by a short-term boost to employment in the sectors that receive the expenditure, e.g. equipment manufacturing or construction. Its longer-term impact on employment is uncertain in two distinct ways. From the firm’s own viewpoint, an expectation is formed during the investment process, and then the practical realization depends on the degree of success of the investment in generating new economic activity that finds favor in the market in the ensuing years – and the degree of success is typically highly uncertain at the time of the investment decision. From the viewpoint of the wider economy, the impact on employment is ambiguous: with investment in labor-saving technology, it may decline even if the investment is successful in the firm’s terms.

5. Why unemployment exists

An implication of this view is that the overall employment available in an economy is the total of that provided by all the investments. In this perspective, unemployment is seen as the gap between the positions created at any one time by firms and the number of people who are available for work, stratified by skill type and location. Thus, unemployment exists because
insufficient investment opportunities are perceived to exist that promise to be profitable with the available combination of workers’ abilities and the necessary wages, as well as other conditions such as the prospective market and the costs of raw materials, fuel and taxes. Its level depends on the perceived scope for profitable investments available to firms within that economy.

Some long-term correspondence of employment and working-age population exists, due to a process of adjustment. If the economic environment includes a tight labor market with high wages, this could be a deterrent to investment, for some firms at least. The converse is not necessarily true, because even with low wages, potentially profitable investments may be limited. The combination of available firm-based resources, including managerial talent and available technology, with the potential demand for the product may not add up to an investment that promises to pay off, even with low wages. The adjustment process is thus asymmetric. (It is also slow, because investments are infrequent, and designed to last for a long time.) The asymmetry means that labor scarcity is typically manifest as a rise in wage level, whereas labor abundance remains a matter of quantity rather than price.

The corollary is that some unemployed workers are likely to be unemployed because no additional investments are seen as possible even with lower wage levels, and even if potential workers would be willing to accept lower wages. As a result, in the aggregate, there are too few jobs to go around. A related question concerns the distribution of jobs: one might think that the workers who happen to be unemployed would be able to exchange their positions with already-employed workers by offering to work for a lower wage. But wage undercutting seldom occurs. Why?

There are three ways that wage undercutting could operate: (i) existing workers could agree to take a pay cut, (ii) unemployed workers could seek to undercut the existing ones in an existing firm, and (iii) a new firm could enter the market, with lower-paid workers.

(i) is unusual, but has been observed e.g. when a firm is in serious trouble, and the alternative is closure or large-scale redundancies.

(ii) could occur (a) because the firm’s previous investment decision was suboptimal, i.e. the jobs that were newly created or maintained could have been offered at a lower wage. Such an error in decision making may occur from time to time, but in general it is probably safe to assume that firms’ decisions at the time of investment are close to optimal, given their managerial capacity and the uncertain nature of information about future economic conditions. Alternatively, (b) there could be a major change after the investment was made, e.g. a shock that resulted in the availability of an unforeseen pool of potential workers with the right skills, in the right location. In such a case, this shock would be the principal cause of the possible undercutting situation, rather than the mere existence of currently unemployed potential workers who are willing to accept lower pay.

(iii) does occur in some industries, as new firms enter the market with a new business model, a more efficient technology, or a more cost-effective location (see Scenarios, below). In such cases, decision making on wages is typically integrated with decision making on broader work practices, as well as with other factors such as technology and location.
Another implication is that the behavior of *individual* potential workers does not impact the number or type of jobs. An unemployed person who is particularly diligent in job search is more likely to achieve employment, but this is merely at the expense of someone else.

In summary, firms are seen as proactive, forward-looking agents that take the initiative. In contrast, (potential) workers are seen as reactive, in the sense that they react to the employment positions created by firms – acknowledging that the decisions which create these positions are arrived at by taking into account the existence and characteristics (number, skill type, going wage, etc) of the *potential* workers. These could currently be unemployed, employed elsewhere, self-employed or inactive. Firms propose; workers react.

This perspective is supported by evidence that workers tend to be backward looking in their reservation wages, possibly driven by perceptions of fairness and/or backward-looking reference points (Akerlof and Yellen, 1990; Falk et al., 2006; Della Vigna et al., 2014; Koenig et al., 2016). The past orientation of workers, together with the setting of approximate wage levels by firms as a strategic decision at the time of investment, intended to last for its whole duration, imply a substantial degree of inertia in wage levels.

6. The broader context

The real investment led economy viewpoint only applies to a real economy dominated by capitalist firms, in the sense of firms that are able to buy in all their inputs, including labor. It is well recognized that unemployment in this sense is a relatively recent phenomenon, since the industrial revolution. Outside the capitalist context, e.g. in peasant agriculture, the equivalent is *underemployment*, which typically occurs when the working-age population is large compared to key resources such as land, rather than to the size of workforce implied by the sum total of investment.

Non-capitalist firms survive in the modern context as self-employed individuals. Confusingly, they are referred to as “entrepreneurs”, leading to widespread conflation with the idea of innovative entrants who bring about creative destruction. Entrepreneurs in the former sense are generally observed to have lower productivity than capitalist firms (e.g. GEM, 2017). It is also a myth that most innovation is due to entrepreneurs in the sense of outsiders (Hsieh and Klenow, 2017).

One implication of the real investment led economy viewpoint is the possibility that firms may have funds available to invest from retained profits, but lack suitable investment possibilities for future production. The possible consequences are that they will invest in property or the financial sector instead of in the real economy (“financialization”), and that if this occurs on a widespread scale, secular stagnation of the real economy could result.

Another implication concerns the nature of investments – they are not all equal in their consequences for the jobs they create. Some investments provide employment with high productivity, good pay, and a degree of job security. Others only create insecure, low-productivity jobs, where the main aim is to minimize the wage bill.

Substantial changes in the labor market of many rich countries have occurred in recent decades. A firm may decide to classify its workers as self employed in order to avoid legal requirements relating to sick pay, taxation, etc., to reduce costs. This may be accompanied by
flexibility in the number of work hours offered to the workers, e.g. zero-hours contracts, a way of fine-tuning the expenditure on labor to the current demand. This situation is created in the course of investment, e.g. in setting up a minicab firm\(^4\) with its office etc., that is geared up for engaging drivers as and when they are needed. There is actually nothing new in this strategy, which was present in the early twentieth century, e.g. in dock work in Britain.

Its most recent form involves the use of software – also a type of investment – such as Uber or Deliveroo. In some of these cases, the authority relationship of the capitalist firm is retained, e.g. in Uber's control over what drivers do, its use of ratings, the requirement to take a certain proportion of clients, etc. However, the situation is somewhat different from the "traditional" form of employment that had (approximately) guaranteed hours, because the variability in the number of available work hours raises the possibility of underemployment.

The type of employment created also has implications for socioeconomic mobility at the societal level. In particular, at a stage in the development of an economy where managerial and professional jobs are created on a hitherto unprecedented scale, the possibilities for upward mobility are transformed. In the real investment led economy perspective, large-scale upward social mobility is a matter of the available positions – rather than of the attributes of the upwardly mobile individuals, the nature of their education, etc. Clearly, both aspects are important in practice.

7. The relationship of the real investment led economy perspective to existing views

This perspective contrasts with standard neoclassical theory in several respects. That theory puts forward models relating to the decision making of (potential) workers, and of firms – respectively the supply of and the demand for labor. Workers choose whether or not to accept employment, based on a comparison of the offered wage with their reservation wage. Firms' decision making is seen as a comparison between employing one more or one fewer worker with the difference this would make to production – respectively marginal cost and marginal benefit – given that the firm already exists, and has an established production system with premises, equipment, etc. Neoclassical theory implies that the forces of demand and supply rapidly bring about an equilibrium in which there is neither excess demand for labor, nor excess supply.

The real investment led perspective holds that it is true that a process of adjustment does occur (see Section 5), but the timescale is far slower than neoclassical theory would imply. It involves investment decision making that depends on the economic environment, including likely future demand – the micro (or meso) analog of Keynesian aggregate demand. In addition, the bulk nature of investment implies that the decision making is "lumpy", rather than the "smooth" process implied by the theory – although very small firms' decisions may approximate to the notion of adding or removing an individual marginal worker. Finally, it is an open question whether investment is best seen as part of an adjustment process leading towards equilibrium, or as a disequilibrium process of creative destruction – a topic that is beyond the scope of the present paper.

---

\(^4\) In Britain, a minicab company is a firm that coordinates the work of self-employed owner drivers. They pay a fee to the company for the services of the call centre, which takes bookings and schedules the work. The firm usually supplies radios as well. Each driver has to wait to be told when a trip is available, and a standard flat-rate charge applies.
In the neoclassical perspective, wages are – or “should be” – flexible, and the labor market “should” clear. The observation that persistent unemployment is quite frequently observed has led to the modification that (nominal) wages are assumed to be sticky downwards. This asymmetry can be compared with the asymmetry in the real investment led economy perspective, which is because some investments do not promise to be profitable even with low wages, so that a shortage of jobs does not lead to a fall in the wage level; in contrast, a shortage of workers does raise wages (section 5).

As its name suggests, the real investment led economy perspective places investment at the center of the economy. A large proportion of economic activity is the result of previous investments by firms. This means that the investment decision, the driving force, is strongly endogenous to this system in a causal sense. This contrasts with the view that attributes labor market phenomena to shocks to e.g. demand or productivity.

The familiar concept of rent attributable to employer-employee matches, divided between the two parties, is represented here by (a) a contribution to firm profitability, and (b) comparison with the worker’s outside options, i.e. when unemployed, inactive or self-employed, or the wage in a previous job (for matches following in-job search), as appropriate.

In the literature, some emphasis has been placed on the distinction between new jobs and existing or continuing jobs. The proposed perspective suggests the need to further distinguish between jobs that are newly-created (strategic) and newly-negotiated (tactical – e.g. new matches). It explains why wages in new matches (tactical) respond to the current unemployment level, whereas those in existing jobs do not (Pissarides, 2009; Koenig et al., 2016) – they were already set during an earlier strategic decision.

The real investment led economy view takes labor supply as given. It does not address participation rates, i.e. flows between unemployment and inactivity, or in-job search as contrasted with the job search of unemployed potential workers. For these aspects of the labor market, a complementary approach such as a search-and-matching model is required. This does have consequences for real-economy investment, via the economic environment. For example, if previously inactive workers become available for work (“unemployed”), e.g. as a result of a policy intervention, or if largescale immigration suddenly occurs, firms can take this into account in their investment decision making for the future.

The real investment led economy view has little to say about how jobs come to an end for an individual worker. This is because the starting point in that situation is an employee who is already in a relationship with an employer, implying a much more symmetrical situation: either party can bring the arrangement to an end. The employee can leave for personal reasons or to move to a better job, or s/he can be sacked for low productivity, indiscipline, etc. But it is highly relevant to larger scale job losses, as discussed in the next section.

8. Scenarios

As previously mentioned, any attempt to explain where jobs come from needs to be applicable to a range of scenarios. They are described in this section, each with an example drawn from real-world experience. There is some overlap between them. The subsequent two sections deal with how these are treated in existing theories and in the real investment led economy perspective, respectively.
(a) Setting up a new company

In recent decades, a number of low-cost “no frills” airlines have been set up. These seek to exploit the high costs of traditional airlines, and the opportunities afforded by deregulation in many jurisdictions. Here we are concerned with the investment process, and its consequences for the labor market. To set up an airline, it is necessary to acquire (buy or lease) aircraft, obtain landing slots and check-in facilities at suitable airports, and obtain government approval for safe operation. All types of staff need to be recruited and trained: aircrew, cabin attendants, ground and maintenance staff, etc. Other costs include fuel contracts, advertising and insurance. All these need to be budgeted for, and finance arranged if not already available. Strategic decisions are necessary on schedules and fares, as well as logo and livery, seating types, meal provision, etc. (See e.g. Creaton, 2004 on the setting up of Ryanair; also The Economist, 2011.) Low costs have been achieved using such policies as quick turnaround times to maximize utilization of fixed capital, the use of cheaper airports, and restricting the air fleet to just one type of aircraft so as to minimize the costs of staff training and the required inventories of spare parts.

One result is the creation of the required number of jobs in each category. A strategic decision needs to be made on the wage level that is deemed necessary to attract, motivate and retain staff of the appropriate skills and quality. This is likely to be strongly influenced by a comparison with similar jobs in similar organizations. Other decisions include policies affecting staff morale, including unionization – different airlines have taken very different directions in this respect. The staff-related decisions are intrinsically bound up with all the other aspects, as in the decision on aircraft types which affects both training costs and maintenance costs.

(b) Introducing a major technological change

Modern technology has made it easy for people to arrange their own bookings, for flights, hotel rooms and myriad other purposes. This has led to major labor market changes. In the hotel industry, for example, a whole stratum of middle management has disappeared as the work of booking rooms has been devolved to customers, as well as being hugely simplified. This process continues to expand, with self-service checking in now being introduced, which has further implications for staffing levels (Worgull, 2017). In addition to the loss of hotel management jobs, some employment is created in software-based firms such as booking.com, but on a much smaller scale.

The loss of employment opportunities here is a by-product of investment decisions of the software-based firms, and the way that hotels react to this new situation. The loss of staff positions clearly entails a large cost saving, but there is no corresponding loss of benefit to the hotel. Neither is there necessarily any deterioration of the service: self-booking is plausibly more convenient for customers, especially as it can be done from a website comparing the availability and prices of different options for any particular location, using a rather small number of clicks. Arguably this is less arduous than having to contact individual hotels by telephone in order to find out if there is a vacancy, and then to make a booking. It also has an added advantage in providing an automatic record and confirmation of the booking.
(c) Relocation of production

Firms may decide to transfer production from one location to another. This can take many forms, e.g. the twentieth century move of much manufacturing within the US, from the traditional industrial regions in the north to areas further south that had lower costs (see below). More recently, it has tended to involve an international relocation, offshoring, especially for large firms.

One example is the decision by Dyson, the British domestic appliance manufacturer. In 2002, vacuum cleaner production was moved from England to Malaysia with the transfer of 800 production jobs. The decision was made on the basis that cheaper labor was available in Malaysia that would not involve a proportionate loss of productivity (Gribben, 2003). The result was an increase in profits and a rise in the number of high-skilled jobs (e.g. in engineering) in England, as well as an expansion into the US market (Gow, 2003).

(d) Plant closure

Plant closures are a major reason for job loss, and their occurrence is strongly cyclical (Davis et al 1998). They often cluster: other plants of the same type in the same industry and country frequently close at around the same time. An example is the closure of integrated steel mills in the US in the 1970s and 1980s. Recessions occurred in 1973-75 and 1981-82, greatly reducing the demand for steel, at a time of rising energy prices following the oil crisis. Other underlying causes included competition from East Asia, and the introduction of a new technology, electric-powered minimills, that competed with integrated mills in low grade steel production. Both these factors gradually increased during the 1970s and 1980s, and the recessions – especially the later one – precipitated massive losses (three billion dollars in 1982) and large-scale plant closures. Output shrank by more than a third, and 150,000 workers were made redundant (Rowe, 2016). There was “marked uncertainty about the profitability of new capital investment” (Davis et al., 1998).

Plant closure is the mirror image of investment. Firms that are in a strong position – perhaps because of low unit costs, or because they have a successful new product – are able to take initiatives that are potentially profitable. They invest, and one result may well be that their weaker competitors are forced onto the defensive, and have to scale back production, or to try and imitate their more successful rivals. This may involve restructuring. When all responses fail, the result is likely to be plant closure, or even the demise of the firm.

(e) An economically depressed region

Formerly known as the Manufacturing Belt, the Rust Belt is a large area in the US from the Great Lakes to the upper Midwest States. Its proximity to iron ore and Appalachian coalfields, and to the Great Lakes and other transportation infrastructure, propelled it to industrial prosperity in the early twentieth century. However, from the middle of the century, it began to decline due to a combination of factors including relocation of production to the southeastern states, automation that reduced the need for labor, and trade liberalization that encouraged offshoring (Crandall, 1993; High, 2003). Manufacturing employment fell by a third by 1996 and has continued to decline since then. The result has been not only a loss of jobs, but also a fall in median household incomes of approximately a quarter in some cities. In addition, there has been large-scale out-migration from the former major industrial centers – Cleveland, Detroit, Buffalo and Pittsburgh lost about 45% of their population. Other consequences
included declining tax revenues, swelling welfare rolls, a poor standard of education, and social problems such as crime and drugs.

(f) A major depression

The story of the Great Depression of the 1930s is well known. The boom years of the 1920s ended with the 1929 New York stock market crash. Many countries were devastated in the depression that followed. Unemployment rose to over 20% in the US (Garraty, 1986; Hamilton, 1987). Some of the job losses were immediate, due to the failure of businesses, but beyond that, the failure of investment meant that future production, and therefore future employment, was compromised. This is not the place to discuss the various theories on the deeper causes of the Great Depression, but it is clear that the proximate cause of the medium-term fall in employment was largely a shortfall in real-economy investment – whether that was in turn due to inadequacy of aggregate demand or of the money supply, to debt deflation, to pessimistic expectations, or some combination of these and other factors.

9. Existing theories and the scenarios

(i) Standard neoclassical theory

In standard theory, there is symmetry between firms and workers: their choices play an equally important role. Workers’ choices could be relevant in the first three of the above scenarios: they can decide whether or not to join a new firm, including a software-based hotel booking firm, or to take employment in a plant that has relocated to their area. However, this could only occur after the investment decision had been made and announced. The potential for it to occur would affect firms’ decision making, but only indirectly, via their perception of the availability and cost of the workers they require. The decision making of workers is clearly irrelevant in the scenarios of a plant closure and a depressed region. In the Great Depression, the suggestion would have to be that workers voluntarily left their jobs, and refused new ones, because the offered wages had become too low, so that they preferred “leisure”. This conflicts with the evidence on what happened in the 1930s, e.g. on the shortage of vacancies. More broadly, the causes of involuntary layoffs, as in plant closures, are totally different from those of voluntary quits: for example, the former increase in recessions, whereas the latter fall sharply (Akerlof et al., 1988; Davis et al., 1998).

Firms’ decisions in the scenarios are also at variance with neoclassical theory. In the case of a new airline, the textbook notion that the firm decides on the flexible factor, labor, to fill an already-established number of slots (e.g. machines) corresponds poorly with actual decision making: the decisions on aircraft, slots, etc entail a corresponding complement of the various types of staff required – the business plan specifies everything simultaneously.

In the case of the hotel staff, there is no calculation of marginal costs and benefits. The elimination of the need to pay managerial staff is not accompanied by a loss of benefit to the firm (nor by a loss of convenience to the customer), so there cannot be a trade-off. The relocation of Dyson could be said to involve a comparison of marginal costs and marginal

---

5 A vast number of theories have attempted to explain unemployment – see e.g. De Vroey [2004]. I have therefore been highly selective, focusing on those that have been highly influential and those that appear to be the most persuasive.
benefit, but this would encompass premises, equipment and supply chains, not just employment.

With a plant closure, or a wider downturn in space or time, the firm is typically faced with losses that cannot be allowed to continue indefinitely, and economic conditions where investment does not promise to be profitable. It is possible to describe this in marginalist terms, e.g. that the losses correspond to the difference between marginal cost and marginal benefit, but this is at the plant level (lumpy), not a question of an individual marginal worker.

(ii) Efficiency wage theories

Various theories have been proposed, that firms pay a premium “efficiency” wage. This could be to discourage quitting – especially when turnover costs are high (Schlicht, 1978), to raise the potential cost of workers shirking (Shapiro and Stiglitz, 1984), to promote a sense of reciprocity and fairness (Akerlof and Yellen, 1990), and/or to attract higher-quality staff (Schlicht, 2005). There is good evidence that some firms do pay efficiency wages to at least some of their employees (e.g. Raff and Summers, 1987), and receive efficiency gains as a result.

In addition, it has been suggested that this higher wage has a role in creating unemployment: it is above the market-clearing level, and therefore the labor markets fail to clear. This puts the focus on the worker’s decision: a comparison of the wage level and her/his reservation wage, the lowest wage at which s/he will accept the job. However, in the scenarios the appropriate comparison is rather the firm’s decision-making process. The decision to close a plant depends on its loss making (or failure to make adequate profit), which is the difference between revenue and costs; the workers’ reservation wages are irrelevant. Similarly, the absence of investment in a new company, technology or location depends on the going wage, not the reservation wage (which is unknowable to a potential investor), together with other potential costs and the likely revenue – the absence of sufficient promise of profit.

The situation is somewhat different for the last two scenarios, a depressed region or a recession, which involve also the aggregate level. In these circumstances, the higher wage would have some protective effect at the macro level, by raising aggregate demand, thereby reducing the degree of contraction. On balance, a higher wage bill could therefore increase employment.

Thus, the efficiency wage theory may accurately describe much of the behavior of firms – especially the better-off ones, or those that have a relatively low wage bill because they are capital intensive. But it is irrelevant to the perceived need to explain the lack of market clearing – a perception that derives from the ingrained assumption that labor markets “should” clear. This creates an apparent need to explain the observation that they frequently do not do so. It leads to the situation where the aim of theory becomes to account for the divergence of reality from standard theory, rather than to explain reality itself.

(iii) Canonical search-and-matching theory

Another theory that attempts to explain non-frictional unemployment, the failure of the labor market to clear as predicted by standard theory, is search-and-matching theory. It suggests that equilibrium is delayed because the equilibration process is slower than traditional theory assumes. This is due to the difficulty that potential employees and workers have in finding
each other – a friction that means we never quite get to equilibrium. It implies that unemployment is just a question of reallocation – “the closeness of the match between the desired and actual characteristics of labor and capital inputs” – involving distance, skills, etc (Davis et al 1998, 106). In this perspective, “unemployment consists of workers who lose their jobs because it is not to their advantage (and to their employer’s advantage) to continue employed” (Pissarides, 2000, xvi), which is a poor account of, for example, plant closure.

Clearly, search-and-matching theory only applies once jobs have already been created, that is, after the investment decision and its announcement. It has nothing to say about firms’ decisions on scale, technology, location, closure, etc, nor about depressed regions or recessions, and is thus irrelevant to the various scenarios.

On the other hand, it is well suited to describing the dynamics of labor supply, and could thus be relevant to the tactical adjustments described above, that occur in the course of production. It therefore has a complementary role to the real investment led economy perspective proposed here.

(iv) John Maynard Keynes

The perspective closest to the real investment led economy viewpoint in the previous literature is that of Keynes in the General Theory:

“The unusual feature of Keynes’s analysis is its recognition that all the power is in the hands of producers. This is not because they occupy a monopsony position in the labour market, but arises simply from the temporal ordering of the process of producing for (uncertain, future) market sale: firms decide how much employment to offer on the basis of their expectations, and in the decentralised system of Western capitalism, if these decisions do not absorb all the labour available, that is just too bad” (Chick, 2007).

The parallels are clear: in both accounts, the firm is the initiator and this gives it a degree of power. The quantity of employment depends on the totality of investment. In addition, both theories involve time. In Keynes’ case, the wages are paid before the product is sold; with the real investment led economy viewpoint, the longer timescale of investment and its medium-term payoff is involved. Money is necessary to both perspectives. Uncertainty is central, because the product may not succeed in the market, which is accentuated in the real investment led economy view more than in that of Keynes, in view of the longer timescale and larger number of variables involved in investment.

The theories are fundamentally distinct, however, because Keynes was here referring to production, not investment. The time delay and the uncertainty refer to the gap between the payment of wages and the receipt of revenue, whereas with investment the intended payback period is spread over a number of years. And as is well known, Keynes’ theory of unemployment was quite different from that proposed here.

10. The real investment led economy explanation of each scenario

As previously stated, according to the real investment led economy view, firms invest when the expectation of future gains exceeds the expectation of the necessary costs by an amount
that generates a satisfactory return on capital. This clearly applies to setting up a new company, such as an airline. The types and quantities of jobs flow from the calculations that lie behind the business plan.

With technological change and relocation, the calculation is that a fundamental change is considered advisable, because they would generate a better investment. The costs of the innovatory investment, including managerial time and effort, are justifiable in view of the expectation of better performance in the medium and longer term. Also, innovative investment may be considered necessary in a dynamic sector, to avoid being displaced by competitors.

Understanding the other three scenarios requires a firm-centered perspective, with profitability at its center, if not necessarily investment as such. As stated above, plant closure is the mirror image of investment, which affects firms that are in a relatively weak position. The decision to discontinue production is taken by the firm, taking the economic conditions into account. This sets the agenda, and any other decisions then have a secondary status, including possible negotiations with workers or their representatives.

Similarly, in a depressed region or a recession, the firm’s profitability is crucial. With weak demand, costs may need to be reduced in the short term, which will often imply loss of jobs. The firm-centered view in this situation is again the mirror image of investment: a firm in a weak position, faced with difficult decisions, needs to focus on the future scale of operation, possibly reducing it so to minimize losses. It is true that a possible response would be to cut wages (and other costs), rather than to reduce the scale of production, and this has occasionally been known to occur. But usually it is the scale of production that is chosen, because the firm has a continuing relationship with its employees that would be impaired by wage cuts – its priority is the morale and therefore the productivity of those workers who remain employed. According to interviews with firms conducted after the early 1990s recession in the north-eastern United States, “resistance to pay reduction comes primarily from employers, not from workers or their representatives, though it is anticipation of negative employee reactions that make employers oppose wage cutting” (Bewley 1999). Note that this response was phrased in terms of firms’ decisions based on their perceptions of the situation, which accords with the perspective that firms take decisions in the light of their economic environment.

11. The real investment led economy view and puzzles in labor economics

*Non-frictional unemployment*

Perhaps the central puzzle of labor economics is the existence and persistence of unemployment, sometimes on a large scale. The real investment led economy viewpoint suggests that the explanatory focus should be on the opportunities for firms to make investments that promise to be profitable and thereby to create employment. There is no assumption that market forces “should” bring about a labor market equilibrium, particularly in the short term. The cause of non-frictional unemployment was outlined in section 5.

---

6 It is unclear how generalizable such findings are to other places and periods as the research has not been done.
Wage stickiness

The real investment led economy perspective proposes that the approximate wage level is set strategically at the time of investment, albeit that the precise wages may then be negotiated when employees are hired, and again when new matches are made, as well as being modified e.g. during annual pay negotiations. The rough level of wages then persists at least until the next investment is made, or until plant closure. Large changes are usually not made, as firms fear that this would destabilize the situation and might well reduce productivity and therefore profit. This reasoning is reinforced by the observation that contracting the volume of production saves not only the wage bill of the dismissed workers but also (in many cases) the other costs involved, such as raw materials and the costs of maintaining premises; in contrast, very large wage reductions would be needed to make significant savings (Bewley 1999) – as in the context of investment, the wage bill is only one item, albeit an important one, in the totality of the firm’s situation.

This would predict that real wages in capitalist firms are rather stable, with only rather minor adjustments being made between major investments, which is in accordance with the usual real-life situation. Wage stickiness is thus a consequence of the theory, not an assumption. Note also that this argument is distinct from the notion that fixed-length contracts reduce wage adjustment – in the real investment led economy view, the same approximate wage level is retained when one worker leaves and is replaced by another with similar skill level, although it may be modified by prevailing market conditions.

The wage flexibility puzzle and the unemployment volatility puzzle

The wage flexibility puzzle is the observation that the cyclical variation in wages is typically very low, whereas standard theory predicts that wage levels should fall in a recession (Pissarides 2009). Instead, employment usually fluctuates more than predicted – the unemployment volatility puzzle (Chodorow-Reich and Karabarbounis, 2013). The real investment led economy view predicts that wages are roughly stable between one investment and the next, because approximate wage levels are part of the firm’s investment decision. This reasoning is reinforced by the broader arguments given above, relating to the effects on morale of wage cuts compared with the scaling down of production, and to the relative savings from the two alternative courses of action.

Evidence on vacancies

According to the real investment led economy perspective, employment is created by firms – but not necessarily as abundantly as would be needed for everyone who is willing to work to be able to find a job. This means that there is typically (although not invariably) a shortage of jobs, of a magnitude that depends on how successful the economy is, in the sense of the firms being able to find potentially profitable investments. The clear prediction is that vacancies are scarce, and this accords well with the evidence: it has long been recognized that reported vacancy rates are low, that there are typically many applicants for each vacancy, and that vacancy durations are typically short. Classic studies include Holzer (1994) for the United States, Beaumont (1978) and Roper (1986; 1988) for the United Kingdom, and van Ours (1989) and van Ours and Ridder (1992) for the Netherlands. It also fits with the observation that job offers made by firms are usually accepted by a large majority of applicants, e.g. 90% (Barron et al., 1986; 1997).
In addition, this perspective fits well with the less-researched observations (what Manning (2003) calls “casual empiricism”) that it is difficult for unemployed workers to find employment, and that typically only a small proportion of firms report difficulties in hiring labor – although there may well be skill shortages due to a mismatch of desired and available skills. Nevertheless, the idea that there can be a shortage of jobs, e.g. in a recession, meets with resistance in some quarters. For example, Rogerson and Shimer (2011): “workers are constrained from working as much as they would like during recessions, perhaps because search frictions prevent them from finding a job”. This is despite the abundant evidence that recessions are characterized by a shortage of vacancies.

**The correlation of employers’ characteristics with wages**

The emphasis on the firm’s central position in making the main decisions on employment and wage levels naturally leads to the prediction that the characteristics of a particular firm will have important consequences for those decisions.

The best-known observation is the employer size-wage effect (ESWE), which has been documented across a large number of countries and across a variety of different types of sector (e.g. Adamczyk, 2015). A firm’s size “now” is largely a consequence of its degree of success in previous periods. It can therefore afford to pay higher wages, and that also means it can attract higher-quality workers, and is better able to retain them than the less successful firms (Brown and Medoff, 1989). Rather than firm size in itself being an explanation of the higher wage, both of these features result from the superior past performance of firms that are now large.7

In addition to ESWE, other characteristics of the firm have been shown to be correlated with wages (Manning, 2003). These include profits, profits per worker and productivity. In particular, inter-industry differences in wages have long been recognized empirically (Slichter, 1950; Krueger and Summers, 1987; Dickens and Katz, 1987; Krueger and Summers, 1988; Katz and Summers, 1989; Akerlof and Yellen, 1990). For example, in sectors with a relatively low proportion of costs that are due to wages, and corresponding higher costs of raw materials/fuel or capital, firms are able to pay higher rates – wages are less salient for them. And even the gender pay gap has an important firm-specific component (Card et al., 2014).

12. Conclusion

The account put forward in this paper is based on a description of the way that firms operate, notably the time order of decisions and events. In a sense the description is obvious, and probably most labor economists recognize these phenomena from everyday life. The issue is that the phenomena are not represented in the dominant theories – but they do make sense when integrated as the firm-centered perspective.

The real investment led economy view promises to be informative in the context of labor economics. Its account of firm decision making, emphasizing the strategic importance of the investment decision, is realistic. The focus on investment means that the theory is long-term, dynamic and forward looking. This perspective is able to explain non-frictional unemployment,

---

7 This does not mean that large firms are necessarily destined to out-perform their competitors in the future – size is a consequence of past performance, not an accurate predictor of future performance.
to provide a good account of a range of possible scenarios, and to shed light on some of the existing puzzles in labor economics.

We are now in a position to answer the questions posed at the beginning of the paper. Who creates jobs? Firms do. When? In the course of investment. Why? To engage in production, in order to make a profit. How? By using their buying power to attract and retain workers, to develop a relationship of authority with them, and thereby to direct their productive forces.

One of the implications is that individual workers’ job-search behavior primarily affects the allocation of particular workers to particular jobs, not the total number of jobs or their characteristics. An important point here is, however, that the economic environment of the firm – crucial in forming its investment decisions – includes the availability of workers with specific skills and other qualities, as well as the going wage rate for employing them. The role of potential workers in these decisions is thus seen as being at one remove, mediated through firm decision making, at the level of a group of prospective employees, and at the time of the firm’s investment decision when firms and potential workers may not yet have met. This implies a departure from a long tradition within labor economics, that seeks to explain the features of (un)employment by referring to workers’ preferences, e.g. their reference wage, and/or their job-search behavior; these latter factors however remain relevant to the tactical adjustment of wages and working conditions to economic conditions as they change, as well as to labor supply decisions, and to who gets which job.

The real investment led economy perspective explains the existence of persistent unemployment. It does not start from the assumption that the labor market should live up to its name – that the wage level “should” adjust, allowing an equilibrium to be reached. If one takes that conventional starting point, together with the observation that unemployment is common in the real world, one needs to explain the discrepancy. This runs the risk of providing an explanation that is totally implausible – e.g. the notion that something as small as “menu costs” can cause large-scale unemployment.

It also does not portray unemployment as either voluntary or involuntary. It is clearly not voluntary, in the sense that people choose not to work when the wage level falls below expectations. But also, it is not well characterized by the term “involuntary”, i.e. that a person is willing to work at the prevailing wage yet is unemployed, because that formulation suggests that to understand unemployment, one should focus on what unemployed people are prepared to do. Rather, it is a simple matter of the number of jobs created by firms, in comparison with the number of people who would like to work, stratified by skill and location.

The presentation of this perspective is not complete: for example, the relationship of the firm-centered view with other important topics, such as the role of government policies, remains to be analyzed. Its macroeconomic implications also remain to be developed, for example the impact of effective demand for firms’ products on their investment decisions. Other macro topics that have historically been considered important, but have not yet been considered in this context, are the possible existence of a natural rate of unemployment and whether this varies over time; and the relationship of the labor market with inflation. These are all left for future research.
References


Author contact: m.joffe@imperial.ac.uk

**SUGGESTED CITATION:**


You may post and read comments on this paper at [https://rwer.wordpress.com/comments-on-rwer-issue-no-83/](https://rwer.wordpress.com/comments-on-rwer-issue-no-83/)
Employment in a just economy
John Komlos [University of Munich, Germany]

Abstract
The political philosopher John Rawls suggested that a just society is one which would be created behind a "veil of ignorance", without knowing where one would end up in the society’s distribution of talent or other attributes. Today’s labor market does not meet this criterion, because risk averse people would not voluntarily enter it at random, being too concerned about ending up among those excluded, i.e., those without full time jobs which in early 2018 in the U.S. was still 10% of the labor force or some 16 million people. Thus, a just labor market would strive for full employment beyond the implications of the natural rate of unemployment. The latter concept is actually misleading, because most economists and commentators in the media equate it with “full employment” and make-believe that 5% unemployment is full employment which means zero unemployment. Consequently, endemic and large-scale underemployment is accepted as an inevitable attribute of the labor market. This is insidious, because the concept assumes that the institutional structure of the labor market is carved in stone. According to Rawlsian principles the aim should be to bring unemployment down to the minimum feasible rate which in the U.S. is most likely around 1.2% – the rate which obtained in 1944, and which probably represents a feasible lower bound. Instead of the prevailing system, the right to work needs to be recognized as a natural right, because the right to life depends upon it. Several ways are proposed to create an inclusive labor market that distributes the available work in a more equitable way than the current one and envisions a just labor market on Rawlsian principles that risk-averse people would be willing to enter at random.

JEL codes J01, J08, J29, J38, J68

Keywords labor market, justice, full employment, natural rate of unemployment, NAIRU, John Rawls

Introduction
To human beings fairness is an essential aspect of life. So is justice. In fact, they are so important that people are willing to sacrifice a lot to fight for it, not infrequently even their lives. “Universal values of fairness” was a main motivation of recent demonstrations as far removed from one another as Tahrir Square and Zuccotti Park, many of which toppled governments (Stiglitz, 2013, p. xxxviii). Indeed, the nature of justice was a main concern of the first philosophers.1 In fact, justice or fairness is so central to humans – and not only to humans, as other primates are similarly inclined – that one can infer that it has evolutionary roots (Brosnan and de Waal, 2003; Heinrich, 2000). The probable reason that the concern for justice is so ubiquitous is that cooperation was crucial for survival; violating the rules and expectations of the community would not have been tolerated and those who violated them no doubt would have been ostracized, thereby lowering their probability of reproducing their characteristics. Thus, evolution favored the propagation of those traits that predisposed human beings to value fairness and it became an integral part of human nature. In sum, “justice is a human virtue” (Schmoller et al., 1894, p. 4).

In spite of the widespread prevalence of this disposition, the concept is not an integral part of economics, even if the founder of the discipline, Adam Smith, forcefully stressed its relevance

---

1 See, for instance, Aristotle’s Nicomachean Ethics.
in The Theory of Moral Sentiments of 1759, and it is alluded to occasionally, for instance, in the context of reciprocity in which workers may reduce their productivity in response to unfair treatment by their employers (Akerlof and Yellen, 1990; Bewley, 1998; Fehr, and Gächter, 2000; Skott, 2005). Nonetheless, in the main, it remains outside of the mainstream’s purview, well behind such concepts as efficiency even though there is no evidence that the latter is more important to us than the former. Moreover, the concept of justice has not been applied to the labor market as a whole, even if the idea of a fair wage does enter into models of wage determination (Blinder and Choi, 1990).2

This modest essay begins to fill this lacuna by using the political philosopher John Rawls’ (1971) influential concept of a just society to labor markets and argues that their current organization does not meet the Rawlsian criterion of justice.3 A Rawlsian labor market conflicts with current views of full employment based on theories associated with NAIRU.4 We also discuss some institutions that would bring us closer to a just labor market in the spirit of Amartya Sen (2009) who emphasized the importance of our becoming a juster society or a more equal society (Atkinson, 2015, p. 301). These ideas dovetail well with the discussion concerning the problem of rampant inequality insofar as a more equal distribution of work would inevitably lead to a more equal distribution of income (Piketty, 2014; Temin, 2017; Komlos, 2016).

Just labor market

Rawls argues that a just society should be our ultimate goal: “Justice is the first virtue of social institutions, as truth is of systems of thought…. Laws and institutions no matter how efficient and well-arranged must be reformed or abolished if they are unjust” (1971, p. 3). His litmus test of a just society is whether one would be willing to enter it at random without knowing anything either about one’s own characteristics or where one would end up in the society’s social order. After all, if one is unwilling to take the chance of entering the society at random, it would not be moral to wish it on others, as the Kantian (1785) principle of universal imperative also implies. In fact, most risk averse people would be unwilling to enter most labor markets today if they did not know their gender, race or where they will end up in the society’s distribution of talents, skills, inherited wealth, IQ, looks, and other attributes valued in the market. In other words, if they had no information on their endowments and “original position” it would be too risky to enter it.

Thus, today’s labor markets are not just according to Rawlsian principles as people would not be willing to enter them at random. People would be too apprehensive about ending up among the excluded, i.e., those without full time jobs, which in the U.S. in January 2018 was still 10% of the labor force, or some 16 million adults.5 This is substantial but it does not even include an additional – roughly 5 million – prime-age adults who dropped out of the labor force.

---

2 The Rawlsian conception of justice has also surfaced in the discussion of basic income (Van der Veen, 1998; Van Parijs, 1991; Van der Veen and Groot, 2000).
4 “In an important sense, the US economy is now at full employment” (Feldstein, 2015).
5 The official underemployment rate (U6) is 8.2%, but this leaves out those who would like to work but have not searched for work recently because their prospects are minimal. Federal Reserve Bank of St. Louis, Economic Research, https://research.stlouisfed.org/fred2/series/U6RATE accessed March 4, 2018.
presumably mostly on account of being frustrated with their chances of succeeding in finding employment. And the burden of underemployment is much greater among minorities: among African-Americans it was still 13.4% in January 2018.

What would a just labor market look like? Rawls argues that one would have to design it behind a “veil of ignorance”, i.e., without knowing anything about one’s endowments or initial position. Otherwise our current standing in the society is likely to sway our judgment. If one would be willing to enter it at random then it is moral to make its rules universally binding on others as well.

Full employment

If people could construct a labor market from scratch under the above specified conditions, it would be undoubtedly such that full employment would prevail. The reason is that ex-ante (in the original position) only through employment can one be sure to be able to sustain life. The selfish survival instinct behind the veil of ignorance then leads inexorably to an organization of a labor market in which there is a constitutional right to employment for all, because then even the unlucky would be certain to be able to satisfy at least their basic needs.

Instead of accepting a certain amount of underemployment as natural, a Rawlsian would acknowledge that there is a natural right to life. Exclusion from work threatens one’s very existence since work is necessary for survival, the right to life practically implies that we need to be guaranteed the right to work. Pope Leo XIII argued similarly in his famous encyclical, Rerum Novarum. Moreover, the United Nation’s Universal Declaration of Human Rights states that, “Everyone has the right to work... and to protection against unemployment” (United Nations, 1948). The spirit of this pronouncement appears in many other international documents including in the French constitution: “Each person has the duty to work and the right to employment.”

However, a Rawlsian full employment would differ substantially from today’s common usage of the term insofar as full employment nowadays is commonly equated with the natural rate of

---

7 This is the U6 rate. The true underemployment rates is no doubt a few percentage points higher. Economic Policy Institute, “Underemployment,” http://www.epi.org/data/#/?subject=underemp&r=* accessed March 3, 2018.
8 This concept bears some similarity to Adam Smith’s use of an an “impartial spectator”, an omniscient ideal being, in The Theory of Moral Sentiments (1759) and in general to the philosophical concept of an “ideal observer”.
10 “The preservation of life is the bound duty of one and all, and to be wanting therein is a crime. It necessarily follows that each one has a natural right to procure what is required in order to live, and the poor can procure that in no other way than by what they can earn through their work” (Leo XIII, 1891, Paragraph 44).
11 From the Preamble to the French Constitution of 1946 which is still in force today. http://www.conseil-constitutionnel.fr/conseil-constitutionnel/root/bank_mm/anglais/cst3.pdf accessed May 13, 2016. See also the work of the International Labour Organization whose stated goal is “to achieve full and productive employment and decent work for all...” (Muqtada 2010, p. iii). These values were already annunciated in the ILO’s Constitution including the Declaration of Philadelphia of 1944 (ILO, 2008). A recent UNDP report asserts that “Universalism... also applies to labour markets—ensuring that everyone has access to decent opportunities for paid employment” (2014, p. 92).
unemployment (Friedman, 1968, p. 8; 1977, p. 458) or with the NAIRU (the non-accelerating inflation rate of unemployment) (Phelps, 1968, 1969; Stiglitz 1997). For instance, at the January 2016 meeting of the American Economic Association in San Francisco, Martin Feldstein declared that “We are essentially at full employment with the overall unemployment rate at 5%” (Feldstein. 2016a).12 His pronouncement is not unusual, rather such slight-of-hand is standard practice: Ben Bernanke made similar statements prior to the financial crisis.13

The misleading nature of the concept is magnified through the media which absorbs economists’ usage and misinforms the public by repeating continually that 5% unemployment is “traditional full employment” (Washington Post, 2014).14 Note, however, that the term “full employment” is usually qualified in mysterious ways either by putting it into quotation marks or by referring to it as “traditional” or “essentially in full employment”. The implications of these qualifiers must elude the average reader. As a consequence, endemic un- and under-employment – which in the U.S. amounts to more than 16 million people – becomes widely accepted as an inevitable characteristic of the labor market and defined away. Thereby 5% unemployment becomes an acceptable equilibrium value and must be tolerated since nothing can be done about it. It is just the way the economy works, and hence we must put up with it. This is insidious inasmuch as the practice encourages policy makers to be complacent about the plight of a substantial segment of the population. Moreover, in the mind of the public full employment means that everyone who would like to work has a job and it is not at all clear that the economists’ usage differs markedly. That is why Noble laureate William Vickrey referred to the natural rate of unemployment as “one of the most vicious euphemisms ever coined” (1992, p. 341).

What is meant, of course, is that using conventional monetary and fiscal policy we are unlikely to be able to attain real full employment15 (Federal Reserve Bank of St. Louis, 2014; Gordon, 1997). However, these estimates of the natural rate are not very accurate,16 and the Fed arbitrarily increases it in times of high unemployment and lowers it when times improve. It has been as high as 6.2% and as low as 4.7%. Inexplicably, as of early 2018, the official unemployment rate (4.2%) has been below the supposed natural rate since March 2017 by as much as 0.5%. And yet, accelerating inflation was nowhere in sight. How can unemployment be less than the level of full employment? That invalidates the concept of NAIRU, or the official data, or both.

13 He said in 2005 that “House prices will slow, maybe stabilize, might slow consumption spending a bit; I don’t think it’s going to drive the economy too far from its full employment path though.” At a time when he spoke of full employment there were 7.6 million people officially unemployed, 5 million wanted a job but were too discouraged to search and 4 million were working part time although they wanted full time work. "Ben Bernanke Was Wrong," YouTube video, posted by “Marcus C. Macellus,” July 22, 2009. https://www.youtube.com/watch?v=9QpD64GUoXw accessed August 30, 2014.
14 According to The Wall Street Journal, "The U.S. economy is at last on the cusp of full employment... The U.S. will be in a state of full employment within the first half of 2016 according to 56% of the economists surveyed” (Zumbrun, 2015). Two years later the same Journal wrote: “Full Employment is finally here, or at least not far away. That’s according to the vast majority of economists surveyed” (Leubsdorf, 2018). Here is another formulation of full employment citing St. Louis Federal Reserve President James Bullard: “the labor market is ’at or possibly well beyond reasonable conceptions of full employment” (Jackson, 2016).
15 The natural rate of unemployment is not a constant. The Federal Reserve raises it slightly when unemployment is high and decreases it when unemployment is low.
16 The 95% confidence interval was estimated on the late-20th-century data to have a spread of 3 percentage points (Staiger et al., 1997, p. 34).
Thus, the tacit assumption of the NAIRU conceptual apparatus is that the institutional structure of the labor market is held unchanged (Colander, 1998; Naude and Nagler, 2015). But Rawls would suggest that ceteris paribus need not be maintained and we should use other means to move toward a just labor market with real full employment. One way to proceed is to consider full employment as the lowest level of unemployment attainable according to the historical record. In the U.S. that was 1.2% recorded in 1944. Presumably, that level of unemployment was not related to insufficient demand for labor but must have depended on the physical or mental health of those few remaining unemployed. Of course, that was a time of war, but the experience does demonstrate clearly the capacity of the economy to create job opportunities and bring unemployment down to negligible levels given the right set of circumstances.

A Rawlsian labor market

From a Rawlsian perspective the current organization of the labor market ought not be considered just because the opportunity to work – like wealth and income – is unevenly distributed across the labor force. In early 2018 in the U.S. about 79% of the labor force works full time, 17% works part time, and 4.2% are officially unemployed. Yet, the official statistics overlooks the hidden unemployment of an additional 9 million people, so the real unemployment rate is closer to 10%. Part of the problem lies in the organization of the labor market: the custom is that adjustments in the fluctuations in demand for labor generally occur mostly by reducing the number of hours worked so that their labor time falls abruptly from 40 hours a week to 20 or zero. Hence, one is either given the opportunity to work roughly 40 hours per week or one is not allowed to work at all even though part-time work is also a possibility. Would anyone “behind a veil of ignorance” design such a rigid system from scratch, a system with so much uncertainty and instability – with working times ranging from 0 to 70 hours per week even in normal times. It would be more reasonable to have the adjustment occur in the number of hours worked so that instead of dismissing workers, the available work would be divided more evenly among those wanting to work. Hence, an institutional framework that would enable work to be distributed more evenly would be a reasonable solution to this quandary.

Thus, in a Rawlsian framework one would not be satisfied with achieving a level of unemployment consistent with NAIRU. Instead, the Rawlsian aim would be to restructure the labor market in such a way that it would generate enough jobs to reduce unemployment to the target rate of 1.2%. One way would be to reduce the number of hours worked in a standard work week by roughly 10%, the underemployment rate. This would mean a reduction of roughly an hour a day from 8 to 7 hours similarly to what happened when the 10-hour day was reduced to 8 hours in the 1930s. Such a work-sharing system would be a

17 This was the case, even though the labor force expanded by 10% during the war. Unemployment was also as low in 1918.
18 The real unemployed includes 5 million part-time workers who would like to work full time (who are counted as half unemployed) and 6.6 million people who are not officially in the labor force but would like to work if the opportunity. The labor force in this calculation is augmented by the 6.6 million people who have not looked for a job recently but would like to work.
19 Some tentative steps in this direction were taken in the 2012 “Job Creation Act”. Such a program works in Germany where total employment has not decreased at all during the Meltdown (Krugman, 2010). The reduction of the workweek in France from 39 to 35 hrs in large firms in the year 2000 is estimated to have reduced unemployment rate by 1.6% by 2002 (Du, Yin and Zhang, 2013).
more equitable shock absorber of a decline in the demand for labor than the current system (Baker, 2011).

Other arrangements that would have ameliorating effects include profit sharing wages in which case wages would increase in good times and decrease in recessions so that workers would not have to be fired, keeping the share of total wages in revenue unchanged (Weitzman, 1984). Encouraging cooperatives would also be useful inasmuch as such firms are more likely to adjust pay to fluctuations in demand rather than the number employed (Craig and Pencavel, 1992; Pencavel, 2002; Rosen et al., 1986). One could also mandate that the government become the employer of last resort, in a similar manner to the government’s role as lender of last resort (Colander, 1981). That way the government would provide stability to the labor market similarly to the backstop it already provides to the financial system (Wray, 1997; Colander, 2009, p. 747ff).

A new institution – comparable to the Federal Reserve’s role in finance – could provide similar stability to the labor market. “If governments can take on the role of lender of last resort, then we should be willing to see government as the employer of last resort” (Atkinson, 2008). Atkinson calls for “The government... to offer guaranteed public employment at the minimum wage to those who seek it” (2015, p. 303). Paul Krugman supports Atkinson’s idea and calls it an “old fashioned idea, but probably a very good one” (The Graduate Center, 2015, @22:42 minutes). Robert Solow concurs: these are “very very useful ideas” (The Graduate Center, 2015, @24:18 minutes). After all, it would create to an inclusive economy in which no one is deprived of the opportunity to work and no one is excluded and stigmatized20 (Stiglitz et al, 2015; Junankar, 2011). As Stiglitz put it: “The rules can and should be rewritten, in ways that promote… employment and reduce exclusion” (2016). After all, during the New Deal the Work Progress Administration hired as many as 7% of the labor force, which in today’s terms would amount to some 10 million people (Margo, 1993, p. 43). That alone would put a considerable dent in the underemployment rate.

Such a system would increase the quality of life, because it would reduce the psychological burden of unemployment, increase leisure time, and reduce envy by reducing conspicuous consumption. In addition, it would be a much fairer method of distributing the pain associated with a diminution in the demand for labor than the prevailing rigid system.

**Rawlsian wages**

Equating wage to marginal product does not meet Rawls’ criteria of justice insofar as much of it is a rent, i.e., return to attributes of the employee that are part of a random allocation. The employee did not do anything for his/her genetic endowment so that any return to characteristics that are in perfectly inelastic supply such as looks, talent, physical size, IQ ought not accrue to the employee and should be taxed at a progressive rate. Rawlsian disposable income should be entirely a function of effort that includes schooling and other investments in human capital. After all, just rewards ought not be based on the luck of a random initial allocation.

---

20 Moreover, in the age of the information technology revolution it ought to be possible to match vacancies to willing workers instantaneously, thereby eliminating frictional unemployment completely.
Conclusion

As the Nobel Prize winning economist Paul Krugman recently put it,

“I am, and I think lots of us, are to some degree Rawlsians; that is, we think that at some level you ought to think about society in terms of what would you want if you didn’t know who you would be behind a veil of ignorance. And you do not have to be a rigorous Rawlsian to feel that it is right, it’s appropriate, to care more about diminishing misery at the bottom than increasing comfort at the top… that is a good thing from the point of view of some notion of justice” (Graduate Center, 2015, @4:00 minutes).

The labor market, as currently constituted, fails the fairness test according to Rawlsian principles. His litmus test of a just society is whether we would create it behind a veil of ignorance, i.e., prior to knowing our personal characteristics and how the market will value them. Insofar as a goodly share of the rewards to labor today depend fundamentally on the luck of birth such as the genetic lottery or the family of birth, Rawls asserts that risk averse creatures as we are, would not dare to enter today’s labor market at random. Hence, it is not a just institution. We ought not to wish on others that which we would not dare to do ourselves.

Hence, Rawls would consider a just labor market one that is at full-employment and not the one in widespread usage today based on theories associated with NAIRU or the natural rate of unemployment. As Stiglitz suggests these concepts merely provide policymakers

“reasons not to attempt to address unemployment…. These ideas provided intellectual comfort to central bankers who didn’t want to do anything about unemployment. But there were strong grounds for skepticism about these ideas… The underlying hypothesis that there is a stable relationship between the unemployment level and the rate of acceleration of inflation has not withstood the test of time…The use of the term ‘natural’ unemployment rate suggests that it is ‘natural’ and natural things are good, or at least unavoidable. Yet there is nothing natural about the high level of unemployment we see today. And these ideas are being used by those that don’t want government to take steps to do anything about it” (2013, pp. 328-329).

In addition, Rawlsians would introduce different shock absorbers into the labor market instead of the rigid system of today. It would be much more reasonable to distribute the burden of shortfall of available work more equitably than concentrating it among some 20 million people in the U.S, as the labor market functions today. If one were designing a labor market from scratch, one would surely construct one that lowered the uncertainty associated with being unemployed. Working less would also increase leisure time available to improve the quality of life for the employed population.

Similar to the wartime experience, effective demand could be increased perhaps not as Paul Krugman suggested – tongue in cheek – by declaring an impending alien invasion (Krugman, 2011), but by improving education, eradicating slums, repairing decaying infrastructure and investing in new ones, reducing pollution, and investing in renewable energy. There is no shortage of productive investments. These projects could create enough jobs to achieve full
employment for many years to come (Brenner and Brenner-Golomb, 2000; Vickrey, 1992; Warner, et al., 2000).

Admittedly this essay is incomplete in the sense that there are multiple other inequities in the labor market. Full employment would not by itself solve the problem of discrimination and privilege in employment systems (who gets what job and who is treated well) and current legal recourse has been insufficient to prevent various types of discrimination (which have to some degree been naturalised by economic theory under the euphemism of endowments). So, fairer also extends to resolving problems in these terms. However, full employment institutions are likely a step in the right direction to achieve other kinds of fairness for work.

"Work is a necessity, part of the meaning of life on this earth, a path to growth, human development and personal fulfilment" (Francis, 2015, paragraph 128). A fairer distribution of work would be important not only to provide the means to making a living but also because underemployment has destabilizing social and political effects (ILO, 2008; Pius XI, 1931). Underemployment generates negative externalities such as an increase in criminality and an increase in stress and anxiety about losing one’s job. Work is important also from a psychological perspective: underemployment is degrading and makes one feel unwanted (Muqtada, 2010; Junankar, 2011; UNDP, 2014). The underemployed do not consider themselves useful members of society and suffer from diminished self-esteem. Their skill depreciates during extended spells of unemployment so that it becomes more difficult for them to find a job. In other words, underemployment increases social misery. For instance, the underemployed are twice as likely to be sad or depressed than the employed and 50% more likely to be angry (Marlar, 2010). They are also more likely to be struggling financially (54%) in contrast to 38% of the employed (Manchin, 2012). This is hardly a negligible matter, especially since endemic real unemployment is likely to be with us for the foreseeable future (Summers, 2014a; 2014b; Brynjolfsson and McAfee, 2012) unless we begin to think creatively about Rawlsian approaches to full employment.

Thus, the concept of the natural rate of unemployment “is an idea that is past its sell-by-date” (Farmer, 2013; Galbraith, 1997). “It is essential that we continue to prioritize the goal of access to steady employment for everyone” (Francis, 2015, paragraph 127). This is particularly important at the current level of inequality as Atkinson asserts: “the present levels of economic inequality are intrinsically inconsistent with the conception of a good society.” A good society or a just society? Either way, it should be clear that we have a long road ahead of us to rethink our economic theories, concepts, and associated policies.

---


22 As the Nobel Prize winning economist Robert Solow put it: “Extreme Inequality is bad for the democratic political process” (Graduate Center, 2015, @3:06 minutes). He also calls it “repulsive” and “immoral.” Even conservative ex-Federal Reserve Chairman Alan Greenspan expressed the opinion that “if you have an increasing sense that the rewards of capitalism are being distributed unjustly, the system will not stand” (Alan Greenspan 2007, @2:36 minutes).
References


Kant, Immanuel, 1785. Grundlegung zur Metaphysik der Sitten.


Author contact: John.Komlos@econhist.vwl.uni-muenchen.de

SUGGESTED CITATION:

You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-83/
Managing the engines of value-creation
J.-C. Spender [Kozminski University, Warsaw, Poland and Rutgers University, NJ, USA]

Caveat

This essay is a rumination on managing private sector firms, why they exist, how they lead to economic value. Many writers see these firms as capitalism’s ‘engines’. The essay explores methods that go beyond presuming managing is or could be a rigorous science. Given our cultural commitment to science and its ability to reveal Nature’s simple rules (such as $e=mc^2$ or $pv=K$), many hope for computable models of firms and managing. This is odd – for history shows we have been researching firms and managing them for centuries without finding either simple rules or evident progress towards them. The essay is not more of this failing program and so may seem puzzling to those expecting science. Nor does it offer seven habits, 11 tips, or any other magic potion. Rather the view that managing is difficult and important and should not be trivialized, especially by those, such as academics, who benefit from value creation by others.

The essay’s first point is that while science’s achievements are indeed amazing, rigorous science is not the only way we try to understand our condition – wherein the economy is central. Economics often pretends to be the science that has taken over our politics ("It's the economy, stupid!"). Science’s popularity has almost displaced religion, theater, poetry, and the other ways we understand ourselves and our doings. It claims thinking is simple and can be imitated by our computing engines.

The essay takes off from Aristotle and turns on subtleties often ignored when discussing thinking. Its focus is on ‘uncertainty’, on ‘knowledge-absences’, and on ‘imagining’ as the way we cope with ‘not-knowing’. We may be able to compute what we know but must imagine what we do not. This was well known to the Ancients – though made famous again by Don Rumsfeld’s remarks about the unk-unks. The essay follows Frank Knight’s intuition that firms ‘exist’ only because of uncertainties and would not exist without them. There is also Ronald Coase’s argument that in the real world all transactions are costly, implying managers must be entrepreneurial, engage uncertainties directly and create economic value sufficient to ‘cover’ these costs. Managers are not helped at all by ‘explanations’ or theories that ignore uncertainty or presume costless transacting. Firms are made up of ‘boundedly rational’ people who, when managed effectively, imagine purposively and so engage complex uncertainties. The practice may create economic value.

The essay is written to managers rather than economists or management theorists. Yes, I borrow ideas from writers I mention. But this is only to help those familiar with our literature see the essay’s construction and flow. The nonacademic reader does not need to pay them much attention. The citations and bibliography are not essential – after all neither firms nor managers were invented by theorists. Yes, the essay is complex in places but is anything about the human condition simple? It moves from a discussion of economic value, to business uncertainties, to the application of entrepreneurial imagination, to managerial practice, and finally onto the process of economic value creation. The analysis eventually hinges on the language entrepreneurs create to shape the activity of others, so I may seem to
offer a ‘language-based’ theory of the firm. Which is fine but misses the crucial dynamic as I try to address Coase’s famous question “Why do firms exist?” I borrow the notion of ‘engine’ from engineering theory, showing how the firm cycles between ‘states’ and the First and Second Laws of Thermodynamics help illuminate its activity.

Note the gender-free terms ‘ne’ (he/she) and ‘hir’ (his/her) are not misprints.

What is economic value?

As The Economist noted “value creation is a corporation’s raison d’être, the ultimate measure by which it is judged” (2009). Corporations are presumed to be the engines of our capitalist system. For economists like Deidre McCloskey capitalism is the remarkable mode of socio-economic order that has harnessed corporations to create massive economic value (wealth), ‘lifting all boats’, and bringing billions of us out of abject poverty and into the modern world. It follows that the business manager’s primary goal is economic value creation. Economic value is not the same as profit (an accounting/legal concept); there can be value without profit and vice versa. Value creation is one of economics’ deepest puzzles and economists have argued about its nature and sources for millennia (Fogarty, 1996). On the one hand price, on the other cost. Use value or exchange value? Objective value versus subjective? And so on. There is no consensus. Yet Joseph Schumpeter, among many others, argued economic value must be the basis of all economic analysis. Economics begins with ‘truck, barter, and exchange’ – creating, trading, and consuming ‘things’ of value.

This essay is about managing, not economics, which has its own problems. Economics is one project, measuring managers by the economic value they create quite another. Beyond profit, the more common managerial measures are ‘goal attainment’ or the creation and maintenance of order, as in ‘command and control’ or the stability and predictability of the group or firm – leading to classic notions of managing like POSDCORB (Planning, Organizing, Staffing, Directing, Coordinating, Reporting, and Budgeting). Clearly ‘good management’ means many different things but none seem directly connected to creating economic value. OK, we assume that long run failure to create value threatens the business’s survival, forgetting liquidation, mergers, and acquisitions are strategies to ‘maximize shareholder value’ – presuming that the most important value to be maximized.

The discussion of economic value, along with much Western thought, was shaped by Aristotle. He argued value was based on ‘need’, inherently personal and subjective, without which there would be no exchange. Given people and their valuations differ, exchange may follow. Aristotle also distinguished exchange value from use value. Historians of economic thought point out that until the modern era economics was part of moral and ethical philosophy. Value was associated with ‘utility for good purposes’ and thus with ‘justice’, as in ‘the just price’. Likewise, in the Islamic trading area that extended from China and SE Asia into Northern Europe, economic activity was governed by feudal power and religious beliefs about value.

As the social and political impact of agricultural and manufacturing increased in post-Reformation Europe, slowly at first but eventually explosively, economics emerged as a distinct discipline. Value was no longer measured in feudal or religious terms, it was secular. William Petty sought a good’s ‘natural value’, what had gone into its production that could be compared with its market price or ‘actual value’. Influenced by Islamic thought, ‘natural value’
led onto the ‘labor theory of value’, famously taken up by Karl Marx. The contrary tradition was ‘utility’ or ‘use value’ – illustrated by John Law’s diamonds/water paradox, the first costly but useless, the second (in the days before bottled water costing more than gasoline) inexpensive but essential. Adam Smith showed demand, subjective need, could be set against supply and its tangible costs. The market ‘solves’ the paradox, prices emerge as the ‘value’ shared, driven by both scarcity and cost. Theorists thus saw true value revealed by the process of exchange. Costs are simple when dealing with trading, the capital invested by the purchaser. But they are more complicated when goods are manufactured or ‘changed’, when costs include apportionment for the land and labor ‘used’; hence the triad capital, land, and labor. The Medievals thought prices ‘fair’ when close to ‘just costs’, and profit non-usurious when close to the ‘natural value added’, reflecting a fair value of the merchant’s labor – milled and bagged flour versus the farmer’s grain.

Despite such complications, neoclassical economists concluded economic values were established/revealed when supply balanced demand, when the market ‘cleared’. This way of analyzing economic value presupposed markets functioned well, for any malfunction would mess up the market’s magical power (Invisible Hand) to reveal true economic value. The efficient market could then be analyzed using marginalist notions. Economics became mathematical; religious, feudal, and even legal aspects of economic value disappeared into the background; rigor claimed the foreground. Many accuse economists of fetishizing ‘perfect markets’ and shifting the analysis from the lived world into a purely imagined one. Many management writers align with these critics, arguing modern economics provides little understanding of the managers who design and run the engines of real-world economic value creation. Assuming perfect markets writes managers and managing out of the analysis – save as the fleshy computers necessary to the perfect markets’ operation, at risk of being automated away. The market not the manager is the locus of action. Economic value is defined as a market phenomenon.

And yet – why managers matter

Most management theorists take the ‘existence’ and ‘nature’ of firms for granted – there they are, all around us, some public, others private, some prospering, some failing; who can question their existence without appearing idiotic? While assuming they exist is fine for reading the financial pages or corporate histories, it is difficult to understand value creation so long as value is a purely market phenomenon. How do firms and their managers fit into the analysis? Coase famously asked this question in a 1937 paper and was eventually awarded a Swedish Riksbank Prize (the economists’ Nobel) for doing so (Coase, 1937). Rather than taking firms for granted, he re-defined ‘the firm’ as an alternative mode of socio-economic organization – alternative to the markets neoclassical economists presumed the ‘proper’ mode of organizing. After Coase, markets were no longer the sole mode to be considered in economic analysis. Instead of claiming ‘in the beginning there were markets’, Coase presumed managers (entrepreneurs) were the foundation of economic action; first, choosing between modes of economic organization; second, contracting for the exchange of factors of production and products; and third, managing ‘the firm’ they created. Economics was not just about the second. Which implied ways of evaluating managers based on economic ideas rather than on POSDCORB criteria. Coase also suggested firms were ‘able to do things markets failed to do’ – connecting firms with ‘market failure’. The firm was an alternative apparatus that could do what markets did not. Note Coase did not claim markets ‘could not’. Rather the opposite, markets and firms were alternatives that, in principle, could handle every
'transaction'. In a specific situation, the manager’s choice of firm as his mode of organizing was purely economic, presumed less costly than the markets available. Managing was ‘economizing’, not POSDCORBing.

Following up Coase’s intuition shifts the analysis away from manager-lite axioms like perfect rationality, perfect information, and optimization and toward managing market imperfection. But a ‘theory of market imperfection’ may not make much sense. Yes, ‘market imperfection theory’ is popular among international trade theorists but does little more than express their surprise that international markets do not meet the neoclassical economists’ pre-conditions – nor do any real markets, of course. Absence does not make for theory. Economists came to see an economy’s history as divided between the time when markets were so imperfect that they did not function properly and the various incommensurate notions of economic value (natural, just, use, labor theory) dominated, and the paradoxes remained; versus the time after the economy’s markets began to work well, when imperfections were less damaging and the economy’s paradoxes and contradictions would be resolved as market processes revealed economic assets’ true value as market prices. Economic value became a market-based characteristic, not inherent in the good or service being invented, manufactured, traded, or consumed. Value became ‘what the market will bear’, with no other basis.

The most familiar market imperfection is ‘market power’. It denies the neoclassical axiom that no economic actor has the power to reshape the market’s activity. Market power may arise in many ways, happenstance (finding a $20 bill), deceit (caveat emptor), thievery (holdup or bribery), legal reallocation (inheriting a real estate company), changes in tax law, etc. – any reallocation of valuable resources by ‘non-market forces’. The resulting heterogeneity is always ‘caused’ as a specific event, so there are no easy routes to a general theory. This contrast between the general and specific ways in which we ‘know’ is at the core of this essay, as it was for Aristotle. In the Nicomachean Ethics (Ch. 6) he explored how the contrast limits our thinking. In the 19th century Windelband re-labeled the distinction ‘nomothetic’ (general) versus ‘ideographic’ (specific) – the terms I shall use to be crystal clear. Theorizing is nomothetic, thinking in generalities, probably time-less. Such thinking stands at some distance from how we recollect experience. Experience is ideographic, situated, embedded in time, having a ‘completeness’ theorizing lacks because it is based on assumptions and simplifications (axioms). Relating experience and theory is deeply problematic – the focus of the philosophy of science. Things are made more complicated because everything we claim to know is ‘held’ in a specific language – for language is inherently nomothetic in that it depends on specific speakers and listeners sharing some understanding, some generalization. It follows that no experience can ever be fully captured in any practical language, what we can say is always at some distance from what we felt, setting up the poet’s challenge to convey emotion. We often fall back to ‘you had to be there to understand what I am saying’.

This is no mere nicety to be dismissed as over-the-top academic nonsense, irrelevant to understanding the ‘real world’. The nomothetic/idiographic distinction is with us everywhere. For instance, it drives a wedge between a manager’s instruction and the action that results, and how that is measured. Managers are thoroughly familiar with slippage between thinking, saying, doing, and evaluating, perhaps as Murphy’s Law. Unfortunately, management theorists tend to ignore Aristotle’s cautions. Many cannot accept the specificity or ideographic nature of market imperfection and presume that monopolistic power arises from something nomothetic such as ‘economies of scale’/’declining average cost’. Aside from confusing the general with the specific, the argument pre-supposes the existence of the firm – the firm must
exist before economies of scale can arise. We cannot explain firms’ existence by claiming that once they exist they thrive by creating market imperfections that advance their interest. No doubt once a firm exists, for whatever reason, market power may be important. Managing is then ‘monopolizing’, rather than economizing or POSDCORBing. Thorstein Veblen argued this long before Coase’s questions and many managers would agree their focus is on monetizing competitive advantage (Veblen, 1965). But there is no theory of the firm here, of how to create firms or acquire monopoly power.

More types of market imperfection arise from the heterogeneity of managers’ ‘knowledge’. First, some writers presume markets are inherently unstable and ‘market clearing equilibrium’ or Pareto optimization is never reached. This severely undercuts the claim that markets are the sole arbiters of value. There is not much theory of why markets take time to clear, though evolutionary language often hides the lack. Clearly instant market clearing renders time irrelevant to the analysis, time’s place in economic affairs is denied. Real events take place in the non-deniable ‘real time’ of experience. Business people say ‘timing is everything’ for good reason. Alfred Marshall, one of the authors of ‘marginalism’ in economics, was greatly interested in time, noting four variations: (a) immediate, when time is so short that managers cannot respond to changes in demand, (b) short-term, when managers can change variable cost factors, (c) medium term, when fixed-cost factors can also be changed, and (d) long-term, when external supply/demand factors dominate beyond managers’ control/influence. Whether supply or demand is the stronger determinant of value varies according to the kind of time, as does the situation’s openness to manager’s influence. In this framing, managing is reallocating resources and adjusting market engagements in the light of market conditions, the firm being an apparatus to hold and allocate resources and to contract with suppliers and customers. Conversely, economic time is defined by managers’ powers of intervention. This makes good sense, but what criteria should managers use to choose between possibilities? So long as their choices are based on market valuations the market remains the apparatus for discovering the values that must determine rational choices. Time’s mattering and the market being imperfect means the relevant valuations arise only after managers have allocated resources, and managed the processes of their consumption. The analysis is ex-post and ideographic rather than nomothetic and, once again, cannot explain why firms exist.

When it comes to knowledge, the information necessary to the market’s proper function may not be available. Friedrich Hayek got good mileage out of this observation, concluding a centralized economy would be bound to fail because its planners could never obtain all the knowledge they needed to run it. He suggested setting the actors free from the planners, to interact in their own interest – free markets. The market then operated as the most powerful knowledge distribution device known to Man – a conjecture with enormous political implications. But since the ‘true’ knowledge only arises ex-post as the market settles, there is no ex-ante route to optimal choice – managing remains central. This gap between the information the actor needs and what the market provides is matched by a second gap Hayek overlooked between the knowledge provided and how the actor/manager absorbed it – reflecting our cognitive limits and biases, recently popularized.

Axiomatizing these limits, the firm can be redefined as an apparatus for acquiring and attaching meaning to economic information that goes beyond humans’ cognitive limits. Herb Simon, another Nobel winner, suggested this but provided no clear theory even as he helped invent AI. Today many are excited by AI and ‘big data’, unaware of these technologies’ own limits, of how such machines can and cannot ‘know’, and thus of how they might change our view of firms and managing. The firm is re-defined as a type of computer to be used when the
market, as Man's most powerful information device, fails to process information to the point markets clear. Managing then comprises knowledge acquisition, managing the knowledge’s meaning, and choosing. Some call this ‘strategizing’. A general theory of strategy is an information-based theory of the firm, its context, and of managing the relationship. Many presume ‘strategizing’ is nomothetic, a theory for all seasons and contexts, disregarding Carl von Clausewitz’s cautions about the idiographic nature of battle and politics.

Before turning to how this line of analysis might clarify real-world managing there is a further important imperfection, ‘externalities’. Taking firms for granted also means taking their ‘boundaries’ for granted. Coase challenged micro economists' assumptions with his four ‘killer questions’, not only “Why do firms exist?”, but also “Why are their boundaries as they are? Why are their internal structures as they are? Why is their performance so varied?” Externalities is a micro economists’ term for the permeability of the boundary between the firm and its markets – which comprises suppliers, customers, competitors, inventors, and so on, a characterization popularized by Michael Porter's 5-forces model. The term is often used to point to costs and benefits escaping the analysis, imperfection arising from some agents' positive power to consume what they have not paid for or, negative, having to bear costs without benefitting. When the firm’s boundaries are permeable it is either because managers have the power to reshape the firm’s markets or conversely, others can affect the value of the firm’s assets. As noted above, the firm must exist before it has power. Managing is then focused on ‘boundary management’; (a) applying the firm’s market power and monetizing that into monopolistic rent-streams (above normal profits) and (b) defending its rent-streams from whose who would undo them. Porter's analysis identified some of these external agents as ‘forces’ that can attack the firm’s citadel and undo its bundle of rents. Strategizing pushes back against such invasion, restoring the firm’s boundaries. But despite his ‘value chain' Porter offered no theory of how the firm’s rent streams arise in the first place, to need defending later.

**Dealing with uncertainty**

The previous section points to ways in which real-world markets fail to make managing irrelevant, the neoclassical economists’ aim. This section gets closer to what managers can contribute to a real economy, especially by creating economic value. Clearly managers create, manipulate, and apply economically-relevant knowledge. Making sense of this requires going beyond generalities and engaging specifics, exploring how the ideographics complement the nomothetics. Imperfection is an economists’ idiographic term for uncertainty, an absence of the certainty framed as predictability. Paradoxically the Aristotelian nomothetic/idiographic distinction shows that uncertainty is the precondition to our knowing and thinking. There is no knowing without uncertainty, the experience of not-knowing. As Simon noted, thinking is meaningless to the omniscient. There is no thinking without uncertainty, specifically the impossibility of generalizing conclusively about any experience or fully knowing the practical implications of a theory. Dogmatism, imagining a knowable world closed to experience, certain ex *assumptio*, denies thinking as most of us understand it. There is no possibility of being wrong. The only mental activity admitted is computation, as well done by computers which do not ‘think’ as we do. Many economists and management theorists pursuing empirically testable theories seem unaware such theories are nomothetic even as empirical tests generate idiographic reports that can never fully capture 'what happened'. Aristotle’s two ways of knowing can never be fully or logically reconciled,
falsification, the claim that a contrary experience defeats a theory, can never be conclusive. Further evidence can always upend the conclusion.

No uncertainty, no thinking, and no choice or real-world economic activity shaped by what agents think. Again, this is no mere philosophical nicety, uncertainty is at the heart of all economic activity even if most economists and management theorists dismiss it. Rational actors only trade when they disagree about some property's value ex-ante, when they see its value as uncertain, when they disagree. The market valuation, the price at which the deal closed, is the economic certainty that emerges ex-post, after the exchange, resolving the uncertainty. Coase embraced uncertainty as the fundamental characteristic of real world economics, pillorying his colleagues for ignoring it and generalizing away from the specifics he insisted were essential to economic analysis. His views were molded by his reaction to Frank Knight's ideas (Knight, 1921; Knight, 1951). Coase later said that Knight was the person who most shaped his thinking. Knight, who won the Walker Award that later morphed into the Nobel, earlier suggested that absent economic uncertainties neither firms nor managers could exist or be explained. Coase concurred, leading onto the idea that 'transaction costs' were those of dealing with Knightian uncertainty as firms arose to complement and resolve the market's imperfections and failures. The rest of the essay explores management's part in making this work in practice.

The section above lists some imperfections, implying a corresponding list of modes of managing, each with a specific relationship to a corresponding type of economic value creation. The list of imperfections and of ways to characterize 'the firm' and manage it is immensely long, because each is an ideographic notion grounded, as Coase pleaded, in 'reality' rather than in simplifying abstractions. Absent all imperfections the economy condenses into a space-time point of Pareto optimality – an economic Archimedean Fulcrum where everything is in a state of perfection, the best of all possible worlds. No further economic activity and no managing. From the Coasian point of view, firms play a complementary role in economics, like the 'last mile' for telecommunications providers, adding needed capabilities to imperfect markets and so 'completing' or 'realizing' the economy – entrepreneurs (whom Daniel Defoe called 'projectors') then 'put it in motion' to yield its economic and social (and political) benefits. Firms are imperfect, not the perfect machines many organization theorists pursue. But the firm's imperfections may be able to 'absorb' and 'resolve' imperfections in the markets they engage.

An Economy = \( f(\text{Imperfect Markets} + \text{Firms}) \).

This relation cannot be read as nomothetic for every concept of 'firm' is ideographic, contingent on and specific to the 'imperfection' identified. Nor is the plus sign 'simple'. The expression's parts are 'apples and oranges', incommensurate, not related logically. Thus, managing is 'complex', the practice of bringing disparate things together through integration and synthesis. The list of imperfections implies various 'economic realities' to be integrated, a plurality somewhat like the plurality of pre-modern ideas about value. Managers' pursuit of economic value, particularly when they intend to monetize it for their shareholders, hinges on grasping the fruits of resolving the economy's imperfections/uncertainties. Any attempt to theorize this is a 'theory of the firm' (ToF).

The simplest ToF is 'arbitrage', Richard Cantillon's 18th century notion of the entrepreneur who, for instance, knows the price of nutmeg is low in the Moluccas and high in Amsterdam and trades on the difference. The uncertainty is specific to geography and commodity
The entrepreneur, not a nomothetic generality, enacts an idiographic ‘project’ like that of Antonio and his colleagues in *The Merchant of Venice*, that ‘completes’ this ‘economy’ with a set of contracts – farming, shipping, selling. The nutmeg is moved from farmer to consumer, perhaps at a handsome profit. The voyage is uncertain, as is nutmeg’s price on landing, to say nothing of principal-agent problems at every stage. Cantillon’s firm is defined as ownership of the project’s contracts and resources, and the owner/entrepreneur’s ability to transfer them on his signature (*firma*). Coase commented that the arbitrageur’s scope would be transformed by technological change, modern logistics, and so on but would still hinge on grasping specific uncertainties and preventing others from doing so. Arbitrage moves resources through space to where they have higher value – a ton of nutmeg in Amsterdam has higher value than in the Moluccas. And through time, a ton of corn may be less valuable at harvest time than in Spring. Acting on such exogenously driven values may make the economy as a whole better off; trade can increase the total value of resources possessed. Whether the entrepreneur benefits thereby will remain a question of distribution, the tenuous boundary between entrepreneur, firm, and market. Ne may miss monetizing a share of the new value resulting from successful integration.

Jean-Baptiste Say’s 19th century entrepreneur is up one level of complexity and uncertainty from Cantillon’s. Say’s entrepreneur ‘moves resources out of an area of low productivity into an area of higher productivity and greater yield’. He goes beyond the arbitrageur’s exploitation of difference in value to exploit differences in productivity which, in turn, requires bringing some resource into economic relationship with some dissimilar resource. Say’s firm transforms resources rather than transfers them. His firm is what ‘the firm’ generally means today, a mode of economic organization distinct from ‘markets’ that transforms inputs into outputs. Markets do not ‘exist’ as firms do. Firms appear to have solidity and identity. In contrast, ‘market’ is the economists’ term-of-art that summarizes multiple idiographic trades/transactions. These transactions may have little in common with each other. Some writers presume markets can be differentiated by the commodity exchanged. It gets confusing because markets also seem to transform one commodity into another – nutmeg into cash into ships and so on – but it is the entrepreneur who does this, not the market.

Economists and management theorists hope a firm’s process adds value. But how can this work? Say’s firm adds transformation to arbitrage, integrating resources, lifting one to higher value by integrating it with others. But the process is ‘complex’, as is the plus sign above. It is no simple addition because real-world ‘resources’ are not simple; they are complex, uncertain, and incompletely understood because understanding them stands on incommensurate axioms. As Penrose argued, there is no certainty about a resource’s use or value. Much economic theorizing presumes homogenous resources, a single commodity, certain and completely measurable, fully own-able with full title. Under which circumstances markets can achieve complete integration, but no new value is created. In the real-world all ownership and property is subject to the political system, there is no full title. Real markets and firms can integrate complex and heterogeneous resources, tangibles and intangibles such as ideas, with both nomothetic and idiographic aspects.

In well-functioning markets the only uncertainty admitted is the agents’ divergent valuing. The markets do not fail; agents interact and close a deal. Cantillon’s mode of integration – arbitrage – is ‘spot’ – offer and deal, inform and agree. There is no labor or transforming resource to be integrated. Management’s role is to price, find a buyer, contract, and deliver, price being sufficient to close the deal. In Say’s firm the resources are complex and ‘divergent in character’. There is no necessary relationship. For instance, we know no necessary
(logical) relation between apples, flour, butter, sugar, salt, and water that makes for a great apple-pie. Recipes differ from theories and formulae precisely because imagination as well as reason must be brought into play. The situation is complex because much is known nomothetically and can be described definitively – as apple, flour, and so on are defined and sometimes standardized. But no recipe can be complete. Its instructions are never fully sufficient to determine the cook’s practice. The gap is idiographic, the recipe’s uncertainties pointing to a space for the cook to experience projecting his imagination into the pie-making process. People are defined by what they imagine and act on rather than by what they know. The recipe offers the cook an opportunity to transform imagination into value in a specific situation. Even when AI enters the kitchen an optimum formula is not likely to emerge because ingredients and consumer tastes vary in ways that defy formulaic (computable) summary. Note that designing and building an apple-pie production line and a logistic system for getting them to consumers merely pushes resolving the uncertainties to the system’s boundaries as issues of quality control and measuring consumer taste. James Thompson likewise distinguished managing the firm’s rational core from managing its boundaries.

Cantillon’s concept of the firm was of trading. Say extended this to include transforming and integrating. Business has been around for millennia and was much discussed in medieval times, so these characterizations were not novel. Rather these writers provided more precise ways of talking about business. Later writers focused in structuring and administering the processes of transformation, integration, and engaging input and output markets. The analysis split into theories of designing and controlling firms (organization theory or OT), theories of persuading people to populate them productively (organization behavior or OB), theories of how to persuade people beyond the firm to sell and buy the necessary goods and services (purchasing and marketing theory) and theories of how to deal with competition (strategy). While the early ideas still underpin the discourse, much has changed – more products, services, manufacturing, trading, data, managing, and more writing. But, ironically, modern analytic methods have become more scientific than those in the past, proposing management as a science, squeezing out of consideration the very uncertainties that Knight and Coase argued were preconditions to firms’ existence. No question the literature generated by management writers since WW2 is a magnificent academic achievement and a platform for today’s global management education industry. But its impact on management practice seems less so.

A typology of uncertainties

There is some heavy stuff in this section – but we cannot get beyond today’s literature on managing as rational decision-making and connect with managers’ practice without engaging uncertainty. All attempts to define uncertainty must fail – by definition, for to define is to take as certain, axiomatic. Those who see uncertainty in terms of probability stand on the certainty of population statistics. Knight saw such modified certainty as ‘risk’. Yes, risk management is important, just as is distinguishing knowing definitively from knowing statistically. But the difference here is methodological and neither mode grasps Knightian uncertainty. Probability is logical/nomothetic, computable. In contrast Knight’s notion was implicitly idiographic, the sense of an absence of certainty arising from an ideographic experience of not-knowing. Something failed, what was expected did not occur – why not? Was the causal sequence (nomothetic) adopted wrong, or did the fault lie with the situation’s ideographic characterization – its initial conditions etc.? Such questions must still be expressed in language, thus standing on what is known. Like us all, Knight struggled with Aristotle’s
nomothetic/idiographic distinction, the failure to relate knowing and experiencing, the inevitable separation between the totality and immediacy of living versus explaining it with abstract concepts.

Knight studied science, religion, and philosophy before switching to economics. He knew the limits to human knowing have been explored for millennia in philosophy, religion, theater, and every other form of the arts, e.g. figure/ground reversal, Velázquez’s *Las Meninas*, or the confusions of *Midsummer Night’s Dream*. One striking medieval metaphor is that ‘it is not given to Man to enter God’s Mind’, to arrive at the Archimedean Fulcrum from where everything is certain and things are what they seem to be. Coase was one of the few who carried Knight’s intuition about uncertainty in economics further. But neither Knight nor Coase wrote much about managing. We must dig deeper. The nomothetic/idiographic distinction points to a state of not-knowing that awaits integrating the parts into a reasoned practice that resolves the distinction. This is the ‘micro foundation’ or ‘micro-institutionalization process’ of value creation. Something similar happens in the natural sciences, leading to the tectonic shifts in thinking Thomas Kuhn called ‘paradigm shifts’. Einstein achieved one by bringing physicists’ notion of mass together with their seemingly unrelated notion of energy – changing both and paving the way for nuclear energy. Until synthesis happens experience keeps reminding us that we know multiple things in multiple ways, none for certain. Our knowledge suffers all manner of imperfections, contradictions, and fragmentation. Crucially, these defects are not in the things we seek to know. These may well be ‘real’ – coherent, logically constructed, existing, simply ‘there’ as ‘realists’ believe. Rather the defects in our knowledge are aspects of how we know. We cannot know anything for certain or objectively. Knowing is subjective, an aspect of us, even when we claim to know facts. Knowledge is a human artifact, an aspect of consciousness. We are the source of all the uncertainties we can be aware of. The effective manager/entrepreneur’s special talent is to dig into these for those that can be engaged with imaginative practice to their benefit – mini-Einsteins of the economy.

The most familiar mode of not-knowing is being ‘ignorant’ of what can be known, a mark of our scientific era. In prior times the most pressing forms of not-knowing were often religious, such as fearing God’s vengeance, unknowable. Ignorance has been brought to the top of our list of uncertainties by our society’s turn towards science as the ‘one true mode of knowing’. Note science sets out presuming everything worth knowing is knowable, independent of our knowing and researching, that there is an unshakeable Truth. The Scientific Method guides us to overcome our ignorance of this Truth. Science-talk has become privileged in our era. The Internet and media show big money can be made informing people about things they believe knowable. Yet we also see ‘fake news’ and hope the less-privileged media talk reports the Truth rather than Falsity.

Dealing with others’ ignorance is not the only or even the most important entrepreneurial opportunity in our polity. In practice, our knowing and not-knowing is vastly more complicated than the notion of objective Truth or its absence allows. Aristotle reminds us ignorance and incommensurability are completely different types of not-knowing. Again, our enthusiasm for science tempts us to collapse the difference, to treat incommensurability as type of ignorance, presuming we can arrive at the Truth by integrating known facts. But, as suggested earlier, this is ironic – it eventually evacuates the idea of human knowing, rendering knowledge irrelevant, rather than moving us towards Truth.
Karl Popper’s falsification is a curious interpretation of the nomothetic/idiographic distinction that presumes the possibility of a scientific (logical) connection between a theory and an experimental finding. Alfred Ayer’s slap-down showed falsification and verification were not asymmetric, one black swan does not disprove the claim ‘all swans are white’. It merely throws the claim into question. Our sense of knowing is ‘irritated’. When the experimental finding is presented as an application of the theory under test, it is tautological; the result cannot but confirm the theory. Popper’s argument was appealing, but all informative experimentation must be knowledge-independent of the theory being tested, in which case the relation between the theory under test (the hypothesis) and the experimental result is problematic, incommensurable, not conclusive. Rather it is a complex kind of not-knowing that calls forth the experimenter’s judgment about ‘what really happened’.

The distinction between incommensurability and ignorance is no philosophical word game. Resolving uncertainty by integrating such fragments into shapeable practice is the entrepreneur’s route to value creation. Enterprising managers must understand/sense uncertainty enough to engage it. They can never control it completely; the outcome cannot be fully predicted. Key is the observation that experienced managers engage ignorance and incommensurability with very different practices. Ignorance of the presumed knowable leads on to ‘research’. Incommensurability calls for discussion, negotiation, reconciliation. Capable managers’ ability to characterize the situation’s uncertainty and thereby choose an appropriate practice seems natural and obvious. Managers are effective when they can ‘read’ the situation, ‘recognize’ uncertainties, and ‘diagnose’ them into categories of practice. Recognition is an act of imagination. Note the difficulty of ‘teaching’ computers to recognize, for they have no imagination and can only ‘match’ the data they ‘know’ in memory with what they know through ‘seeing’, even when aided by algorithms that speed the memory search. Academics who admit only one type of uncertainty, ignorance, blind themselves to managers’ diagnostic and entrepreneurial skills.

Again, ignorance presumes the prior existence of what is knowable, a coherent and logically constructed ‘objective reality’ that exists already and is unaffected by our research practice. It is as if humankind is characterized as a Single Supreme Scientist, probing Nature with unambiguous yes/no questions. Such science admits no incommensurabilities, everything is presumed equally real and explainable. But the Scientific Method is not our only method of dealing with not-knowing. Analyzing managing changes completely when we admit incommensurability, the fragmented nature of our imperfect knowing, as distinct from ignorance of the perfectly known. We address incommensurability by debating alternative ‘knowns’. Note how people and their subjectivity are drawn in. The ‘negotiation’ process is not ‘objective’ because it hinges on the participants’ particulars, their specific not-knowing. There is no general model – implying there can be no scientific model of managing or ‘the firm’ if negotiation is its fundamental process. Yes, some propose rigorous theories of negotiation but must set out by defining (axiomatizing) the participants and their choosing behaviors, claiming to know people, to have a true theory of the individual. Poets know better. The human individual is not knowable to us; we do not know ourselves, let alone anyone else. Our imagination indicates what we do not know. Consequently, real interaction is more complex, we ask advice, we reflect, allowing some dialectical interplay between alternatives. The practice of dealing with incommensurability, the going back and forth between possibilities, is very different from dealing with ignorance and the scientific back and forth between hypothesis and evidence. Yet both processes are dynamic, implying the analysis of value creation must stand on dynamic models/ideas.
Managers must diagnose before instructing action, just as physicians must determine the patient’s condition before prescribing professionally. Is this an A-situation or a B-situation? Managers’ (and physicians’) diagnostic skills vary widely. Even with the best science, gathering data and reducing ignorance seldom leads to definitive doubt-free conclusions. Some uncertainty remains to be ‘diagnosed’. Good diagnosticians are highly regarded for good reason; their choices often push the boundaries of professional practice – famously in the case of puerperal fever. We might debate ‘internally’, but differences in characterization may arise ‘externally’, such as agent C and agent D disagreeing over value. Here a third mode of uncertainty appears as ‘indeterminacy’, the difficulty of knowing how D will respond to C’s move. Note time and expectation enter the analysis. In circumstances of indeterminacy the manager’s choice of mindful practice is often ‘negotiation’. Rather than researching (scientifically) a presumed reality, managers proceed by bringing contrasting but different ideas, knowns, and doubts together. Entrepreneurship and value creation is more often driven by synthesizing practice than by scientific research. This essay concludes by proposing ‘the firm’ as a managed complex of human interaction that grasps targeted idiographic uncertainties evident in the socio-economy. The resulting practice may lead to economic value creation.

To conclude this section, the three types of uncertainty noted – ignorance, incommensurability, indeterminacy – must be complemented by a fourth – ‘irrelevance’. All human knowledge is held in language. Negotiation requires sharing language. Likewise, the entrepreneur must create a language that enables hir to consummate the complex or bundle of contracts that bring the business into economic existence, no longer just an idea beyond the real world, rather made ready-to-hand to be ‘set in motion’ as Schumpeter suggested. Managers need a language specific to the firm that enables them to issue directions and evaluate the consequences of the motions they generate. There is no nomothetic (universal or formal) language. Even when this language is idiographic and identified/constructed, it may not relate adequately to the ideas and actions necessary for the firm to succeed, and so be irrelevant. One downside of using consultants is that their language, embodied in their ‘strategic tools’, may prove irrelevant to the resources and practices necessary for their client’s success. Likewise business meetings, often considered a superfluous part of corporate life, are often crucial loci for adapting, updating, and promulgating new business language. Again, negotiation is often the most effective route to improvement, calling for managers to engage in skilled listening and persuasion. But sometimes new language does not lead to improvement.

The next section moves towards the practice of engaging uncertainty. The section above claims some grasp of specific uncertainties is necessary to understanding a firm and managing it. Science and its methods are focused on engaging ignorance of the general, one mode of value creation. But managers need also to be adept with incommensurability, indeterminacy, and irrelevance, and consequently with shaping, motivating, and empowering the practices of others. Even in our technology-penetrated era, science seldom drives business success. Schumpeter insisted it was the business application of science that shifts the economy from creative destruction to economic growth, not science’s progress alone.

Creating economic value

Value-creation hinges on engaging specific imperfections/uncertainties. The immediate question is “Which of the several imperfections noted above offers the most illuminating
model of managing?” The answer is probably “Them all” for they are inseparable, each implying the other in the same kind of plurality as began the essay’s discussion of value. Rather than being synthesized into Truth, the entrepreneurial responses to the imperfections are brought together in a coherent ‘actionable’ language, idiographic, contingent to the situation. The ‘nature of the Coasian firm’ is ultimately persuasive language. Edith Penrose argued resources have no economic value absent what people can say about them; a tool is nothing without communicating the skill to use it. Her management team’s knowledge is generated collaboratively through discussions of practice, reconceptualization of the use-value of ‘resources’. Her firm’s growth is constrained by the growth of the language necessary to integrate its bundle of resources and put them in motion. The language must be purposive, oriented towards specific strategic goals. Experience cannot distill into useful knowledge without purposive language. All science-talk is directed towards institutional ends, it is not ‘objective’ or free of social inflexion. Likewise, tacit knowledge is irrelevant without the purposive knowledge that articulates it into the real world, so suggesting the practical results as a pragmatic truth-criterion.

It is easy to confuse the claim that language is core to the firm’s nature with the claim that ‘knowledge’ is a firm’s most important asset, implying there is some meta-language about bringing knowledge to bear on the firm. The firm’s language is all-encompassing. It defines everything that can be known about the firm. It synthesizes the firm’s knowing, thinking, and practices. Yes, the entrepreneur must create a language that holds his idea before it can lead to value. But the managers’ instructions, like a recipe, are never entirely sufficient to the employees’ or contractors’ practice, never fully determining. Uncertainties remain and the individual actor’s imagination must be brought into play. The firm’s language does not evolve on a tabula rasa, it must push previous language/s aside, persuading employees and others to ‘get with the entrepreneur’s program’ rather than pursue prior personal ends. The manager’s overarching task is to control the firm’s language in this contested space. It shapes what others attend to or ignore among the complexities of every real situation. Language is the manager’s instrument to shape how others engage the situation’s ignorances, incommensurabilities, and indeterminacies and grasp them in the pursuit of value. Managing is ultimately a rhetorical practice, a talking game. Aristotle, whose book on rhetoric established its study and which remains supreme today described (rather than defined) rhetoric as idiographic – the development of the ‘most effective means of persuasion in a particular situation’ – contingent on aim, audience, context, history, timing, resources, etc.

Every ToF stands on an entrepreneurial idea that implemented might (a) enrich the economy as a whole and (b) provide the entrepreneur an opportunity to get a ‘piece of the action’. Arbitrage leads to a ToF wherein the purchaser allows the entrepreneur a share of the price paid, perhaps knowingly perhaps unwittingly. The more uncertainties there are about the deal, the greater the entrepreneur’s opportunities to gain from engaging them. The knowledge-absences that separate sellers and buyers and make arbitrage possible are ‘out there in the market’ where ‘effectual’ entrepreneurs can poke around and find them as ‘opportunities’. In contrast, the firm is a closed environment of idiographic language (jargon/corporate-speak) wherein the entrepreneur can ‘hide’ the process of integration that resolves his chosen knowledge-absences. NDAs (non-disclosure agreements), like patents and trademarks, act to keep the firm’s secret out of the market. Part of managing is boundary management – keeping the firm’s integration process secret, the opposite of ‘transparency’. Veblen noted the relationship between secrecy and profit; ‘full disclosure’ eliminates the firm’s opportunity.
Arbitrage is the minimal value-creation process – the firm’s ‘micro-foundation’ – two agents, a negotiation language, an agreement/contract, and a process of contract execution. The connection between one agent’s desire to supply and the other’s demand is ‘simple’; its dimensions reflecting the agents’ diverging views. Value is subjective, changed as the deal’s execution shifts both parties’ view/valuation of the assets they possess. We go to eBay to dispose of the pot grandma kept treasured on the mantelpiece. Now there is room for something else, we have the cash and a buyer has the pot. There can be no deal without the differing views and cash enough to cover the transaction costs. Just as in the practice of relating nomothetic and idiographic knowing, there is dynamic motion at the heart of negotiation – offer and counter-offer until the deal closes. The deal’s uncertainties are never fully resolved *ex ante* nor does the economy halt at equilibrium. It is forever in motion, as Schumpeter argued. The process may be oscillation, flip-flopping between nomothetic and idiographic. Popper’s method of scientific discovery flip-flops between hypothesis and empirical test, and back again, never arriving at Truth. There is an analogy with von Neumann architecture computers – flip-flopping between transferring information and executing computation.

Say’s firm is complex, more connections, some incommensurate, others interactive and indeterminate. There may be resource-transforming labor, perhaps capital equipment too, sometimes regarded as ‘stored labor’. One mode of value creation is extractive. The 18th Century French Physiocrats lived in an economy dominated by agriculture and argued all value is extracted from Nature. Farmers, miners, or well-drillers ‘arbitrage’ with Nature, knowing how to capture and reshape her gifts and draw them into the socio-economy (sometimes, but not always, to her benefit). Adam Smith counted some Physiocrats as colleagues but countered their arguments by pointing to the division of labor and the mechanism by which the human imagination is monetized through improvements in coordination. Though bounded by ‘the extent of the market’ the infinitude of human imagination is set apart from Nature’s ‘materiality’. Smith wrote about manufacturing, but the notion extends to services and other non-market processes. The resources available are raised to new value by modes of coordination that are complex precisely because they cannot be reduced to a logical rigorous model. Smith’s firm’s dependence on imagination makes it an order of magnitude more ‘complex’. There is ‘art’ in imagining how incommensurate resources can be drawn together. Smith showed managing is an art-form, perhaps capitalism’s most fundamental art-form. Knight argued for business as an art-form in a seldom noted paper (Knight, 1923).

The imagination’s place in arbitrage can be framed as occupying the ‘opportunity space’ between A’s position and B’s. The space ‘between’ what is known can be illustrated by Adam Smith’s firm. He defined enterprise as a dynamic process within an opportunity space marked out by capital, land, and labor. The boundary to the space is not general, it is specific to what is known – the quantities of capital, land and labor – but also to the incommensurability of their measurement. How much is this land worth? A contingent market-dependent question, likewise the relation between capital and labor; relations only known *ex post* after the relevant market clears. Smith wrestled with ‘entrepreneurship’ as a fourth ‘factor of production’. This helped confuse science-oriented researchers who set out presuming entrepreneurship is a kind of talent or set of traits that can be measured nomothetically and distinguished from the idiographic event of its application, defined in a way that does not stand on the idiographic knowns being brought together in the instance of applying imagination. The entrepreneurial capacity is capitalism’s version of John Keats’s ‘negative capability’, the poet’s capacity to
enter and occupy the universal space of life’s uncertainties – made familiar as the entrepreneur’s capacity to live with (inhabit) uncertainty and ambiguity.

Today’s managers must bring together more factors of production than Smith noted. In addition to the details of the firm’s supply and demand markets, there are non-market factors such as government regulation, corporate law, and social norms. Every entrepreneur’s appetite for uncertainty is limited, too big a bite, no language can be generated to grasp it. Too small a bite, insufficient value created to overcome the unavoidable transaction costs. Porter pointed to the competition generated by other firms and technological change. There is the task of motivating the personnel involved. Communication issues attract attention. Notwithstanding these many possibilities empirical research suggests that the entrepreneurial aspects of most business activity can be framed within a dozen or so ‘knowns’ – axioms to the firm’s language (Spender, 1989). These knowns or constraints to practice are specific to the firm, there is no general model. Bankers and engineers do not attend to the same matters as psychiatrists. Characterizing a firm’s language cannot squeeze out all the uncertainty. Some remains, Murphy’s Law can always kick in. There is no business model that can dictate how to create value in the booming, buzzing confusion of the real economy. There, as Knight suggested, no uncertainty, no profit.

Figure 1 The entrepreneurial opportunity space (Spender, 2014:179)

The characterization provides a model of entrepreneurial management, somewhat nomothetic but ultimately relating selected knowns with selected unknowns, the selection being ‘strategizing’ (see Figure 1). It has been labeled a recipe. It can be illustrated as the multi-faceted boundary between what is known about the firm’s resources before they are integrated into the firm’s practice and what is known afterwards from experience. The uncertainties can be expressed as unknowns ‘trapped’ between ‘knowns’ that others may know but do not know how to bring into relation – just as arbitrage requires knowing price differences and how to negotiate. The opportunity space cannot be occupied by rigorous calculation; the recipe is not a formula. It is more a place of practice awaiting a rhetorician, an artist of persuasive language, rather than a painter.

The rhetor’s task is to construct ‘propaganda’, the firm’s own ideographic language using what his/his audience already understands or can be persuaded into. It is complex for it must be
‘dozen-faceted’ and embrace all the relations between what the firm knows and what others beyond it know. This language is as close as an analyst can get to the entrepreneurial practice that synthesizes the firm’s knowledge into practice, putting the firm’s resources ‘in motion’. Rhetors know that ‘natural’ or ‘informal’ language can carry judgment and emotion. Formal, rigorously constructed language such as computer code or mathematics cannot. A computer’s only concept of uncertainty is ignorance. It brings it to a halt, awaiting further data or instruction. Everything a computer can know is ‘of a piece’, coherent, expressible in a single language – such as C++. People speak natural language that admits the imperfections identified above as in markets – contradictions, errors, lacunae, gaps with experience, and so on. The downside is that natural language disables ‘proof’ – falsification fails. The upside is that it can capture how we deploy imagination to engage the world’s uncertainties – and collaborate to pursue our goals.

Concluding comments

There is no economics without value creation; economics is not merely about value distribution through markets imperfect or otherwise. This thought is not original. As noted earlier, Coase suggested positive transaction costs are an inevitable and ineradicable feature of the real world. But the less remarked corollary is that there can be no economic activity that does not create new value sufficient to cover these costs. Thus, Coase called for a theory of economic value creation. Again, he was not original in this. Earlier French economists such as Jules Dupuit related economics to thermodynamics (Ekelund and Hébert, 1999). Contrary to what many neoclassical economists believe, economics cannot be a physics because non-zero transaction costs mean it must escape the First Law of Thermodynamics. Likewise, Georgescu-Roegen argued economics is illuminated by the Second Law of Thermodynamics, that there is no real-world activity without energy loss – just as Coase claimed there is no economic activity without transaction costs. Real-world activity bridges different states of being. Engines, like those that move our automobiles, create motion by burning fuel – fuel and motion are wildly incommensurate states of energy, so how to bring them together? Diesel engines cycle – heat (burning fuel) in, pressure, expansion motion, heat out (exhaust), compression motion, heat in, and so on. The cycle is a dialog/negotiation between incommensurate energy states; fuel and motion. (Electric cars do likewise with a very different cycle.)

The essay argues the engine metaphor is equally apt for understanding firms, they are value-creating engines negotiating between divergent agents and divergent states of valuation. As with the Second Law, non-zero transaction costs mean value creation leads inevitably to social costs and externalities. Whether these exceed the economic value created is always determined by politics. Thus, there can be no non-political theory of the firm or of managing it. Management is a political activity.

My overarching claim is that managing is practical philosophizing about the property-based world we inhabit. We cannot avoid grappling with how people know, how they do not, and how they apply imagination to generate value-creating practice. Academics who assert managing can/should be scientific and rigorous are hopelessly out to lunch and have nothing pertinent to say to managers. Economic value is only created as the total value of the resources available to the economy is increased by overcoming uncertainty. Sometimes the process is simple, as is proclaiming conch-shells found on the beach are currency. Economic value is always an aspect of a specific society – and its politics. Note there was commerce in
the death camps with values unrelated to the world beyond the wire. Economic value is arbitrary, a human/social artifact. It has no objective foundation. Politics always determines value, just as the 500 rupee note was lately proclaimed worthless or drug prices go up when possession is made illegal. Managers always operate within an idiographic political-economic space and look at it through the lens of the relevant recipe/business model, looking for answers to the eternal managerial question “What does it mean for us?”

The essay claims ‘the firm’ is best grasped as a rhetorical process that transforms others’ reasoning and imagining into practices that lead those engaged to re-value their assets – whether tangible or intangible. The firm’s own language is the closest an analyst can ever get to the idiographic nature of ‘the firm’ as comprised of these value-creating practices. Note how mission statements are important less for what they say than how the practice of creating them reshapes the firm’s own language. The language’s explanatory capacity can be assessed by its closeness to how those inhabiting the firm talk. The gap between many economists’ and management theorists’ rigorous science-talk and such firm-specific talk is generally considerable.

Economics is not society’s only way of looking at and understanding itself. In the past religion was dominant, with its own recipes for maintaining status and effecting change. Many thought reason and science would provide more clarity, and today economics seems to dominate. But it cannot be science warmed-over, with claims of market-driven certainty. To the contrary, economic activity only happens when it engages the uncertainties of our circumstances. Rather than a rigorous science the real-world economics Coase sought would be capitalism’s poetry, its politics by other means.

References


Author contact: jcspender@icloud.com

SUGGESTED CITATION:

You may post and read comments on this paper at [https://rwer.wordpress.com/comments-on-rwer-issue-no-83/](https://rwer.wordpress.com/comments-on-rwer-issue-no-83/)
The effect of academic business studies in Germany and America in the modern era
Robert R. Locke  [University of Hawaii, USA]

J.-C. Spender, a leading authority in management theory and management education, after previewing this paper, noted that it clarifies for the first time that there are two models in business studies “(A) German management and German management education, which differs substantially from (B) the Anglo American model now considered to be the ‘one true way.’” Both models have their origins in educational events that occurred over a hundred years ago: in Germany beginning in 1898 in the Commercial Business School (Handelshochschule) created in Leipzig, expanded by 1910 to five more (Berlin, Mannheim, Munich, Frankfurt on the Main, and Cologne), then by 1920 to two more (Nuremberg and Königsberg), to be multiplied thereafter; in America the Wharton business school (reputedly, though not, the first) in 1881, and then a host of others affiliated with private and public universities founded in the next twenty years to be expanded significantly in subsequent decades. Of the two, Spender continues, “(A) is the ‘elder statesman’ of Western management theorizing and education.” Since,” Spender concludes, “(B) cannot be remedied by more of the same (from within itself), it is high time that (B) folks study the (A) tradition” (email to author, 9 January 2018).

In this article I clarify how the two models came to differ from each other by mid-20th century, primarily because of how they related to the study of economics (Volkswirtschaftslehre in German) in each country, and how this different relationship subsequently led to the disfunctionality of business school education in American democracy, while a different relationship between economics and business economics in Germany inherited from the early decades of the 20th century well served the long-term interests of the nation’s economy and democracy after World War II.

Economics and academic business school education in the U.S. 1900-1970

A good place to begin to clarify this relationship in America is with the Conflict about Methods (Methodenstreit), i.e. which methods should be used in the study of economics, a conflict that erupted in late 19th century Germany. Carl Menger of the Austrian school attacked Gustav von Schmoller of the German Historical School in the former’s book (Menger, 1883). Menger thought the best method of studying economics was through reason that produced general theories applying to broad areas.

The discussion among economists merged into a more general investigation into the philosophy of knowledge. The neoKantian Wilhelm Windelband, as the 19th century closed, argued that science (Wissenschaft) could be divided into two main branches, sciences of the spirit (Geisteswissenschaften) and sciences of nature (Naturwissenschaften). He distinguished between the two in terms of their aim and method. Whereas Naturwissenschaften, he specified, use a ‘nomothetic’ or generalizing method, since they seek to discover law-like and general relationships and properties, social or cultural sciences
(Geisteswissenschaften) employ an ideographic or individualizing procedure, since they are interested in the non-recurring events in reality and the particular or unique aspects of any phenomenon. Wilhelm Dilthey carried the debate further; he contrasted Naturwissenschaften and Geisteswissenschaften in terms of their subject matter – positing a fundamental distinction between the realms of ‘nature’ and of ‘human spirit’ – with each sector the prerogative of a separate category of sciences.

In the Methodenstreit among economists, neoclassical economists sought to identify their discipline as a nomothetic natural science. Stanley Jevons, inventor with Karl Menger of marginal-utility analysis, described their scientific goal: “all branches and divisions of economic science must be pervaded by certain general principles. It is to the investigation of such principles – to the tracing out of the mechanics of self-interest and utility, that [economics] is devoted” (Jevons quoted in Fullbrook, 2007, 162).

Fullbrook describes how the architects of nomothetic neo-classical economics matched their new discipline isomorphically with Newtonian mechanics:

In neoclassical economics, ‘bodies’ translates ‘individuals’ or agents; ‘motions’ translates ‘exchange of goods;’ ‘forces’ translates ‘desires’ or ‘preferences,’ which when summed become ‘supply and demand;’ ‘mechanical equilibrium’ becomes ‘market equilibrium,’ this being when the difference between supply and demand is zero, and ‘physical systems’ translates ‘markets.’…All exchanges were said to magically take place at the prices that equated demand and supply (Fullbrook, 2007, p. 162-163).

No viable natural science meant much unless it could be expressed mathematically. That was the glory of Newtonian mechanics. In the nineteenth century, mathematicians started to work on the social sciences. In 1854 George Boole (1815-64) in The Laws of Thought used mathematics to analyze the logic of language, and Léon Walras mathematized neoclassical economics. With this achievement, he stated in his Elements of Pure Economics, economics had become a “science, which resembles the physico-mathematical sciences in every respect” (Walras quoted in Fullbrook, 2007, 163).

Max Weber refused to accept the claim that disciplines dealing with historical constellations (the ideographic or individuating procedure of the Geisteswissenschaften) are generically different from the natural sciences, even though the latter try to deal with recurrent events and discover general laws or regularities of high probability. In his view, cultural studies are distinctive only in that they originate in the investigator’s sense of what is culturally significant. But once a question has been accepted as significant, it is necessary to formulate concepts that will present the relevant evidence “systematically and in greater unity than has ever existed in the actual course of development;” these “ideal types” can then be employed as reference points for the analysis of behavior since they seek to discover law-like and general relationships and properties. Weber, through Talcott Parsons, greatly influenced sociology in the U.S., but not nomothetic neoclassical economics (Cummings et al., 2017, chapter 4, “The Birth of Organization Science: Or What We Could Learn from Max Weber”).

The discussion among economists spread to the U.S., where the newly founded business schools were locating in universities. University economists seriously questioned business schools’ scientific intent. The Harvard Business School, founded in 1908, taking its cue from the Harvard Law School, adopted a case method of instruction. Students read and discussed
résumés of actual business cases, designed by the HBS faculty to give them a taste for real business problems. At Wharton the business professors found most of their curriculum in the business world. Stephen A Sass, Wharton's centennial historian (Sass, 1982), noted that in the interwar period, the neoclassical economists had been cool to the practical thrust of Wharton's business programs and had little interest in the managerial arts and science that were taught in those parts of the school (270-71). Traditional academic subjects were included in Wharton's early business school curriculum (chemistry, physics, etc.), but, inasmuch as the courses were not developed with business problems in mind, these subjects had no obvious relevance to the subject matter taught in the practical courses (accounting, finance, sales, etc.). The business program itself, therefore, for a long time, reflected the gap between economics as nomothetic or ideographic sciences and business school undergraduate education as vocational practice.

The objections to the vocational orientation came from historical and institutional as well as neoclassical economists. Thorstein Veblen, in The Higher Learning in America: a Memorandum on the Conduct of Universities by Business Men (1918) feared the deleterious effects that the creation of business schools would have on educational values in higher education. In this he and other historical and institutional economists (Weber, Schmoller, etc.) differed little from nomothetic neoclassical economics.

The knowledge disparity between nomothetic economics and business school vocationalism increased in the midcentury as economists' confidence in the prescriptive value of scientific methods grew. The commission that the Chicago businessman Alfred Cowles founded in 1932 made important contributions to the development of econometrics. The Second World War dramatically intensified intercourse between science and operations management. Initially Americans learned from British scientists and engineers who began work in 1936 on the “operational use of radar.” The experts realized that they “needed unbiased scientific assessments of the efficacy of radar” and “how to implement the systems in service” (Locke, 1996, 26-28). The success of British wartime operational research is legendary; more important for this story, the British passed on much of what they learned to the Americans when they entered the war. The Americans, with their superior resources, surged ahead. Examples of science gone to war are legion: the economists Stacy May and Robert Nathan applying statistical analysis to war production programs; F.L. Hitchcock elaborating transportation theory to deal with the complicated problem of moving vast amounts of men and material efficiently. The Cold War perpetuated this scientific trend. The management teams the Pentagon organized in order to maximize the input-output ratios for the new immensely complicated and expensive weapon systems looked to science for solutions. Post World War II defense generated significant management applications, among them linear programming, which George B. Danzig and his associates at the Rand Corporation developed for the Air Force in 1947.

Despite Walras’ claim to have turned neoclassical economics into a discipline, which “resembles the physio-mathematical sciences in every respect,” the analytical toolkit they developed in the 19th century proved to be of limited prescriptive value to policy makers and corporate managers. Postwar military planners and the economists who worked with them at Rand believed the new toolkit that operations research provided would finally enable them to transform neoclassical economics into a prescriptive science. At Rand in 1948, the economist Kenneth Arrow used the toolkit in his work on Rational Choice Theory. His book Social Choice and Individual Value (1951) was the “first real classic” on what “is now taken as a given in economics and has spread out into many neighboring disciplines” (Bellah, 2000, 7).
The neoclassical economists Joseph Dorfman, Paul Samuelson, and Robert Solow applied linear programming to their subject as well (in *Linear Programming and Economic Analysis*, 1958). In 1954, Kenneth Arrow and Gerard Debreu announced that they had achieved a mathematical solution of general equilibrium, “the theoretical core of neo-classical economics,” which Edward Fullbrook states “has become the central showpiece of academic economics ever since” (Fullbrook, 2007, 165; Arrow and Debreu, 1954).

The great reforms of business schools of the 1960s and 1970s were brought about by these neoclassical economists who sought to end the reign, in elite schools anyway, of “unimaginative, non-theoretical faculties teaching from descriptive practice-oriented texts to classes of second-rate vocationally-minded students.”(Quoted in Jeuck, 1973, 284) Mathematics requirements were raised, new scientifically qualified staff employed, in order to introduce “the analytic, normative, mathematical, and scientific mode of instruction” (Jeuck, 287). I describe the introduction of the new paradigm in operations research (Locke, 1989, Chapter I). But the best treatment of how the new paradigm entered U.S. business schools is Chapter 6 “Disciplining the Business School Faculty: The Impact of the Foundations,” 233-290, in Rakesh Khurana’s 2007 book (Princeton University Press).

The triumph of nomothetic thinking in neoclassical economics, when it entered the reformed business schools, wiped out vocationalism – in the interests of prestigious “fundamental” nomothetic research and teaching. Ironically the triumph also eliminated the historical economists and institutionalists, who had themselves denigrated business school vocationalism, from economics departments and the reformed business schools’ faculties. It also led to the creation of elite business schools where the professors and their MBA students, armed with the new prescriptive techniques, formed the ranks of a group of symbolic analysts who, in the era of financialization of the world economy, command high salaries and bonus incomes. Khurana’s study of Harvard Business School MBAs shows the elite school trend. He cites a survey of first jobs for graduating HBS students: between 1965 and 1985, students’ entry into financial services and consulting “rose from 23 percent to 52 percent” of graduates (Khurana, 2007, 328–29). The same shift happened in “other elite schools, such as Wharton and the business schools at Stanford and the University of Chicago.” By 2005 “among the 180 principals and managing directors in the 20 largest investment firms, 73 … [held] an MBA from one of the six elite schools (Harvard 51, Chicago 7, Columbia 6, Stanford 5, Dartmouth’s Tuck 3, and Northwestern 1” (p. 349).

The German model (A), Spender’s “elder statesman of Western management theorizing and education”, 1900-1945

If vocationalism long prohibited the creation of a focused discipline in U.S. business schools, in Germany academic tradition required it. From the beginning, business studies professors in the new *Handelshochschulen* (HHS) had not only sought to define the content, scale, and scope of a new discipline, but to acquire all the academic accoutrements (research seminars, Ph.D. research and *Habilitationsschift* programs necessary at the tertiary (*Hochschule*) level to respond to the German idea of knowledge (*Wissenschaft*), as dynamic not static, driven by research.

At first they did not use the term business economics. Some referred to the new studies as commercial (*Handelslehre*), others to the development of a *Privatwirtschaftslehre*, as opposed to economics (*Volkswirtschaftslehre*). Professor Schmalenbach in Cologne and his followers
did everything they could to extend their studies into industry (to broaden the new discipline from commerce to industry, to make it a Betriebswirtschaftslehre (BWL) that was not limited to commerce (Handel), and he, with the support of men like Nicklisch in the Berlin HHS, imposed his view. “Schmalenbach’s contribution”, Alfred Issac wrote in 1922, “is to have helped the factory take its rightful place in business studies” (Issac, 1922). By 1938 business economists held nineteen chairs in technische Hochschulen, from which they developed a new subject in management studies, the Diplom Wirtschafts-Ingenieur (Economics Engineer). Prion, after he joined the faculty in the Technische Hochschule Charlottenburg (Berlin) introduced the degree in 1924, from where the program subsequently spread to other technische Hochschulen and technical universities. Dipl. Wirt.-Ing. degree holders from engineering schools were and are especially prized as agents in consultancies working with industrial firms.

Concurrently, they sought to create the academic accoutrements that would permit them to define and develop a discipline. The first generation of professors did not have BWL degrees because the field of study had not existed in academia. Nor were the commercial schools able to promote the new field because they were not empowered to grant research doctorates. But faculties of business economics when incorporated into new civic universities at Cologne and Frankfort on the Main did. Eugen Schmalenbach, in the first, and Fritz Schmidt, in the second, seized on the opportunity to train the second generation of BWL professors in the field they were developing and to feed them into academic chairs in the BWL educational sector.

Inevitably, business economists and economists would differ about the validity of a new discipline; they did in another Methodenstreit that broke out in 1912. The economists M. Weyermann and H. Schönitz (whose 1912 study provoked the conflict), claimed that business economics had two faces, one as science (Wissenschaft), where knowledge is an end in itself (Selbstzweck) that scientists promote through disinterested work, the other as a technical study of all the possible economic prescriptions and knowhow imposed from the outside, i.e., from nomothetic or ideographic science or from the firm (e.g., to maximize a firm’s performance). Eugene Schmalenbach in Cologne argued against the science outlook (Wissenschaft). He famously claimed that BWL was a technology oriented Kunstlehre, a discipline of praxis-related knowhow, whose task is to present practical applicable knowledge, which attempts to answer questions about what are the optimal means to employ (Aim-_means relations) in business and industry. Accordingly, a Kunstlehre in its presentation is concerned first of all with overcoming the discrepancy between the given and the desired situation in firms; secondly, it is concerned with the efficiency of the means employed.

Accordingly, what made the Methodenstreit in economics different from that in business economics (BWL) was the view of the latter as a technological Kunstlehre. The business economists in their Handelshochschulen had to answer to the needs that business and industry posed from the outside. Willi Prion, who earned his doctorate in Cologne, made this perfectly clear in the midst of the BWL Methodenstreit in a letter to Schmalenbach published in the Zeitschrift für Handelswissenschaftliche Forschung (vol. 3, 1912-13, pages 231-42), the oldest academic business journal in Germany, founded by Schmalenbach in 1906 and still published with Schmalenbach’s name on the title page. Prion chastised Schmalenbach for misleading people with polemics against the scientific outlook (Wissenschaft):

All who have anything to do with Handelshochschulen know that all [of them], in Cologne, as in Munich, Berlin, Leipzig or Mannheim, attempt exactly the
same way to promote one end: to make the businessman (Kaufmann) effective in business, to show him how a business is set up and conducted, how the organization of the accounts, of the business or of the plant is done, how financing is achieved in commercial life, what role the balance sheet plays for business and for the general public, how purchases, sales and calculations are made in various businesses, what role bills of exchange, checks, stocks play in commercial and economic life, how the conditions of existence of the individual firm, their legal constitution, their effect on economic life in general, etc., take place. All the business schools teach that to their students and all teach it, in principle, in the same way (letter excerpts quoted in Locke, 1984, 2006, 157-58, my translation).

There was not much chance that professors or students would become disinterested scientists in these schools. All agreed that business economics, BWL, had to be a technology whatever its scientific aspirations, that is, to answer to and reflect the needs of business and industry posed from outside Handelshochschulen.

Although Weyermann and Schönitz were willing to concede that BWL had a scientific face, most economists whether in the nomothetic or the ideographic camp believed that economics had to be academic not practice driven, and their opposition to business economics as a Kunstlehre was far from passive. The issue became for university economists one of purely professional Interessenpolitik as they set out to block the penetration of business oriented economic studies into their sanctuary. In 1909 Gustav von Schmoller objected to the appointment of an industry connected, nonacademically qualified (according to university lights) candidate to an economics professorship in Leipzig University, and raised the matter before the already famous Verein für Sozialpolitik in its Vienna meeting; Lujo Brentano and Max Weber followed suit at the meeting of the Association of German University Professors. Both groups denounced this attempt to appoint a Tendenzprofessor (purpose professor) in Leipzig. And the attempt did not succeed (Locke, 1984, 119-21).

German economists’ opposition to BWL as a Kunstlehre, did not have the same effect on the development of business studies in German academia that the aggressive activist agenda of nomothetic neoclassical economists and mathematical modelers had on the hapless American business schools during their reform in the 1960s and 1970s. German economists, especially of the dominant historical school, were not actively working in business and industry in the economy they described; their knowledge was not functionally useful to people in praxis. BWL professors, on the other hand, were activists par excellence, who sought in their interaction with business and industrial firms and through the education of their students to drive their agenda. In The End of the Practical Man, I describe the creation of this new academic discipline up to 1940 in three chapters: “German Business Economics: The Theoretical Achievement”; “German Business Economics: The Educational Achievement”; and “German Business Economics: The Institutionalization of Management” (Locke, 1984, 2006, 155-294).

That agenda differed from the American in one important respect. German BWL drew a distinction between studies of use to management (Lehre für Management) and studies of management (Lehre von Management). Unlike U.S. business schools, with MBA programs that welcomed post-experience students into their ranks and developed executive management programs, German BWL students learned little about ‘management’ as a generic subject. Faculties of business economics did not offer the MBA, and German firms
were not interested in hiring them. One could learn specialist skills in schools (even then there was much to be learned about them on the job), but how to manage was not one of them; it had to be learned in the firm. As Peter Lawrence explained (Lawrence, 1980), Germans did not appreciate “the general processes of communication, decision-making, coordination, and control,” the sort of managerialism taught in American business schools.

The firm-centered development of German academic BWL reflected the outlook of practicing managers, who, much like military leaders, who welcomed the educational system needed to develop the logistics, technology, and the esprit de corps of the armed forces, understood, in a world of bounded rationality filled with unknowns, a Lehre von Management could not be one of them.

This hardly meant that the development of a Lehre für management had to be narrowly focused, and to understand why, when it was being created in the first decades of the 20th century, German ideas about knowhow and knowledge have to be taken into consideration.

The late Ian Glover observed that “in Anglophone countries, two cultures, the arts and sciences, are recognized” (Glover, 2013). But a pejorative distinction is made in both cultures between “fundamental” research and applied science, that is, for example, physics is studied in the prestige universities as a “fundamental” science, engineering treated as an applied science for the less brilliant and gifted. Nomothetic neoclassical economics in the two-culture environment assumes the mantle in the elite universities and business schools of “fundamental” research.

Glover went on to note that [in Germany] rather than two cultures there are three: arts (Kunst), science (Wissenschaft), and Technik, which he defines as “the many engineering and other making and doing subjects, representing practical knowhow (Können),” but also including by the late 19th century scientific knowledge (Wissen). Glover’s point is that we cannot find Technik in Anglosaxonia, because it is not part of their approach to knowledge. Mine is that the culture of Technik provided a broader milieu in which engineers and business economists in late 19th and 20th century Germany could integrate the work world than that provided by reformed U.S. business school education. Nothing illustrates this better than the way German engineers and business economists reached out to each other after the turn of the century.

Professors of business economics worked in the German Commission for Technical Education (Deutscher Ausschuss für Technisches Schulwesen) that added business economics to engineering studies during the first decade of the 20th century. They worked in specialized associations of German engineers (German Mechanical Engineers, German Electrical Engineers, German Chemical Engineers) and others (German Association of Graduates of technische Hochschulen, German Association of Graduates in Business Economics) on projects too numerous to mention here (see Locke, 1984, pp. 269-282). They also became heavily involved in the work of the Frankfort Society for Efficiency Training (Gesellschaft für wirtschaftliche Ausbildung, e.V. zu Frankfurt am Main), a private organization founded in 1903 that brought leading professors of engineering and business economics and government officials together with industrialists and businessmen for the purpose of promoting performance “…where the question of good administration in the broadest sense of the word plays a role in state, municipal, and private enterprise” (Locke, 1984, p. 270).
The Frankfort Society became intimately associated with the new business faculties. Its library and archives were in fact amalgamated with those of the Frankfort School of Commerce in 1911, thereby providing better holdings for members of both institutions. The Frankfort Society also supported the annual meetings, begun in 1907, of professors of business economics teaching in technische Hochschulen (whose numbers were increasing rapidly); in 1908 it directed industry’s attention to cost accounting by conducting an essay competition on the subject. The best essays, which described actual costing systems, were published in Schmalenbach’s journal. The Society sponsored ad hoc conferences, which alerted businessmen and industrialists to problems in business economics. Schmalenbach sought out the engineers as partners, in the sense that he wanted to teach nontechnical administrative skills (accounting, merchandizing, sales) to people in industrial as well as commercial firms. He was successful. In my research, I discovered Schmalenbach and the beginnings of academic German business studies, through references to his work in German engineering periodicals.

The business economists also worked in the Reich Trust for Efficiency (Reichskuratorium für Wirtschaftlichkeit, RKW). Founded in 1922, in the political turmoil of the Weimar Republic, the RKW functioned much like the Frankfort Society, although on a much greater scale, bringing organizations and educational agencies together in a common endeavor that sought to inform the interested public about the best way for a defeated Germany to accomplish economic efficiency. One member of the Trust wrote:

Just ten years after its foundation there were about 4,000 co-workers from employer and employee, industrial and commercial circles, from various branches of science and government, who, with 50 regular employees, bore the principal burden of its 150 committees and work groups, dispersed all over Germany (Büttner, 1973, 11).

The Trust’s committees and work groups were usually headed by prominent businessmen. C. F. von Siemens, for instance, was the RKW’s first president; its second, Karl Köttgen, was a member of the Siemens-Schuckertwerke AG board of directors, and Eduard Mösler, who presided over the Committee on Administrative Efficiency, was a member of the board of a bank, the Diskontgesellschaft AG. But the real work of the RKW was in the hands of the business directors of the committees, subcommittees, and research institutions.

Between 1922 and 1926 engineers monopolized the work. Sometimes RKW committees were composed solely of engineers. Then in 1925 the Trust created the Committee on Administrative Efficiency that reflected the growing stature of business economics in Germany’s mental capital mix. Eleven of this committee’s members were professors of business economics (Tiburtius, Seyffert, Schmidt, Schmalenbach, Rössle, Nicklisch, Mahlberg, Kalveram, Hummel, Geldmacher, and Bucerius). In addition, business economists with academic connections (Bötheme, Müser, and Klinger of the Association of German Business School Graduates, Eicke and Schlüter, assistant professors in business schools) were on the panel.

This committee’s work encompassed various aspects of business administration, including accounting, general administration, sales and distribution, finance, and office management. The transformation it brought to the work of the RKW can be shown in the realm of cost accounting after it was removed from the jurisdiction of the engineering-dominated Committee on Administrative Efficiency and transferred to a subcommittee chaired by Schmalenbach.
Under his guidance the subcommittee’s investigations broadened into commercial and business as well as industrial cost accounting (RKW, *Jahresberichte* – 1931).

Subsequently the RKW added other committees to handle commercial and monetary aspects of the efficiency movement, for example, on banking and distribution, and it also made formerly independent business research institutions part of its purview, for instance, the Research Center for Commerce in Berlin, the Institute for Research on Consumer Goods in Cologne, and the Institute for Research on the Marketing of Finished Products in Nuremberg. Prominent business economists were no less involved in the work of these committees. Schmalenbach’s student Alfred Meier became business director of the Schmalenbach-chaired subcommittee on cost accounting; Ruberg, a member of the subcommittee on merchandizing, and Rössle, a member of the RKW board in Bavaria. Hirsch, Tibertius, and Kühn successively ran the Research Center for Commerce in Berlin; Schäfer, another of Schmalenbach’s students, directed the Institute for Research on Marketing of Finished Products in Nuremberg; Bucerius was scientific advisor in another RKW-affiliated institution, the German Small Trade and Artisan Institute.

Schmalenbach’s subcommittee developed standardized charts of accounts, published in 1927. These *Kontenrahmen* contained flow charts, which “…helped in representing not only accounting plans but also organizational relationships, that is, the flow of paper work, production flows, and the like.” Käfer stated that Schmalenbach’s work “was the basis for all later German uniform charts and influenced considerably the national charts of accounts adopted in other European countries” (Käfer, 1967, p. 73). Moreover Schmalenbach’s 1927 Chart of Accounts became the framework in which subcommittees of the RKW expedited work on the industrial accounting systems of specific industries. Schmalenbach’s committee published, in the RKW monograph series, uniform accounting schemes for the lignite, mechanical construction, iron foundry, weaving, inland shipping, brewing, drop forge, and brick making industries.

The business economists also pushed the management techniques of comparative firm analysis. The RKW, prompted by business economists who had worked on the subject, was especially intrigued by inter-firm comparisons within industrial branches. These ‘branch investigations’ did not look at all the firms within an industry but concentrated on a few, selected according to certain objective criteria (size, location, organization structure, etc.). They became yardsticks against which other firms could measure their industrial performance. The RKW then increased the scope of comparative analysis. Initially comparisons had been made between industrial firms; then, Hirsch, in the Berlin Research Center for Commerce started to do comparative analysis for commercial firms. Bredt and Birnbaum, supported by the Association of German Wholesale Cloth Merchants and the RKW, carried out a comparative firm analysis in the textile wholesale trade.

By the end of 1932 the RKW had completed investigations on a number of industrial and commercial branches plus several branches of retail and wholesale trade in agriculture. The results were not published, in order to encourage firms that feared publicity to agree to investigation, but they were communicated to all firms within an industry, including in some cases those not studied, in order to guide them towards better management practice.

The broad contacts that engineers and business economists enjoyed through their participation in the culture of *Technik*, also promoted the vertical integration of society. The German engineering association – *Verein Deutscher Ingenieure* (VDI) – founded in 1856 had
a lot to do with it, because, although it supported the development of trade schools into *technische Hochschulen* in order to meet the scientific demands of the Second Industrial Revolution (1850-1930), it consistently pitched a large tent. If professors from *technische Hochschulen* headed the VDI, it always included in its membership craftsmen, machinists, and graduates from sub-university engineering schools. Moreover, the technical institutes were institutionally tied to the technical trade school system primarily because they educated their teachers. *Diplom-Ingenieure* formed the sub-university trade school teaching staffs, reaching down to the German apprenticeship system, since entry into a sub-university technical school required the completion of an apprenticeship program.

The *Handelshochschulen* had an even closer formal relationship with nontechnical commercial schools through the degree they established of *Diplom-Handelslehrer*. Candidates for the degree received the same education in business economics as those aspiring to become *Dipl.-Kaufleute* plus additional instruction in pedagogy, for those who held the commercial teacher’s diploma from a *Handelshochschule* had the exclusive right to teach in sub-university commercial schools. Through the *Diplom-Handelslehrer*, academic business economics established a permanent relationship with the sub-university level commercial schools and a link, through them, as *Diplom-Ingenieure* had within the technical trade schools, to the practical world of nontechnical commercial training, because entry into the sub-university sector of commercial education usually required the incoming student to have completed an apprenticeship training program. BWL, thereby, became part of a great interconnected nontechnical educational arc, from *Hochschulen* to apprenticeship training, within the context of the third culture of *Technik*.

Nonetheless it is important not to exaggerate the presence of the culture of *Technik* in Germany in the first half of the last century. It existed in a society run by elites that did not accept cooperative cultures if they meant the elite had to share power with subordinate orders. Neither, when the century began, the emperor nor the aristocrat that ran the German army had any intention of giving up their power. Nor, when the imperial power collapsed in the throes of a lost war, could the cooperative culture that the engineers and business economists shaped, assert itself in the financial and economic crises of the Weimar Republic or the totalitarian culture of the National Socialism dictatorship that succeeded Weimar, embedded in the regime’s concept of the *Führerprinzip*. The firm-centered tradition in German business economics just described, and the nomothetic science that ensconced itself in U.S. business schools are important here for what they tell us as traditions about problem solving in our world.

**Nomothetic economics and the retention of firm-centered, praxis relevant business studies in Germany after 1945**

Germans call 1945 Zero Hour (*Stunde Null*). Under military occupation, the universities closed. Although they had much to learn from the victorious Americans about management education and management, unlike the British, the French, and other Europeans, the Germans had their own traditions in business economics, which spilled over into the postwar period.

They spilled over first of all in human form. The membership of the Schmalenbach Society, founded in 1932 in the name of the premier professor of business economics, shows the continuous work of prominent economists across the chasm of 1945. Among the first five
postwar presidents of the Schmalenbach Society were Willy Minz, Walter Krähe, and Fritz Hardach, all of whom, after studying with Schmalenbach in Cologne, worked with the RKW in the inter-war period. The nine members of the Society’s board of directors in 1960 counted, besides these three, Löffler, Hax, and Potthof, that is, men who had participated in the work of the Schmalenbach Society before the war. Furthermore, the membership directory in 1960 lists Banse, Beste, le Coutre, Cordes, Eich, Hasenach, Kosiol, Meier, Mellerowicz, Münstermann, Rosgowski, Schäfer, and Thiess, men who had contributed a lot prewar to the development of German business economics.

Nor was their 1960 membership in the Schmalenbach Society simply an honorary distinction granted to a discarded generation in recognition of prior service. These men emerged from the war, leading in their domain. Walter Cordes survived the dissolution of the United Steel Trust to become director of a steelwork and a member of the board of director of the August-Thyssen-Hütte; Fritz Hardach had returned from the war to become a member of the directorate of the Hütten- und Bergwerk AG in Essen; Walter Krähe became business director of the Ruhrkohle-Treuhand-Gesellschaft in Essen; Albert Meier, member of the board of the Treuhand Seminar Vereinigung AG in Frankfort, was business head in the consultancy Coopers & Lyland GmbH in the same city; Willy Minz became a member of the board of the Rheinisch-Westfälische Treuhand AG in Cologne; Erich Potthoff joined the board of the Zentralverband deutscher Konsumgenossenschaften in Hamburg; Albert Meier, Auffermann, and Potthoff, in addition to other engagements, worked as independent business consultants. Others reasserted themselves in academia. Hax and Banse followed Schmidt and Kalveram in their chairs at Frankfort on the Main; Theodor Beste moved from a position behind the Iron Curtain to one in Cologne. Hasenack, after working on the dismantlement of Ruhr industries immediately after the war, became a professor of business economics in Göttingen. Münstermann held a similar chair at Mainz University, and Mellerowicz, in Prion’s footsteps, became professor of business economics in the Technical University, Charlottenburg (Berlin). Le Coutre held a business economics professorship in Mannheim. Schäfer remained in Nuremberg as professor of business economics and director of the Institut für Wirtschaftsbeobachtung, where he had done so much before the war to further rational management in the retail and wholesale trades.

The Cologne centered group, as important as it was, only tells part of the story, for it leaves out people whose education in business economics took place outside Cologne and who worked in regions elsewhere in Germany. This includes the most famous man in the postwar scenario, Ludwig Erhard, Konrad Adenauer’s minister of economics, architect of Germany’s postwar economic ‘miracle,’ who had worked on the rationalization of German business in the Institut für Wirtschaftsbeobachtung, an affiliate of the RKW, under Schäfer.

The presence of these men kept the BWL tradition alive. Some among them actively cultivated it. Professor Wilhelm Hasenack, who held a chair in BWL in Göttingen after 1950 and had earned his doctorate under Schmalenbach in Cologne and then finished his Habilitationsschrift under Prion in the Technische Hochschule Charlottenburg, wrote a series of biographical death notices published in the Zeitschrift für betriebswissenschaftliche Forschung (Research Journal in Business Economics) about the accomplishment of the prewar generation. Within their bailiwick, the research followed the lines of specialist prewar BWL since nothing in their inherited studies involved nomothetic model building.

But Germany had lost control of its destiny; the Americans with their nomothetic ideas about economics and management were now in charge. Those who had opposed Schmalenbach’s
view that BWL was a *Kunstlehre*, revived the *Methodenstreit*, in order to turn BWL into a science (*Wissenschaft*). The chief architect of this view, Erich Gutenberg, ally of the German neoclassical economist Erich Schneider, in his very influential 1951 book pushed for the adoption of the *mathematisch-deduktive Methode* in BWL, as the Americans were doing in the reform of their business schools at the time and proselytizing all over the world.

However, there were several obstacles in Germany, aside from the continued presence of the prewar generation in education and praxis, that tempered the adoption of a nomothetic outlook in BWL. Some were the consequences of defeat and occupation. Germans were forbidden to have a computer industry; they were also forbidden to work in atomic energy, and they lacked a powerful, aggressive military establishment that could sponsor think tanks, like the Rand Corporation, which worked on the decision models that so impressed U.S. neoclassical economists after the war. And in addition they also did not have the symbiotic linkups between private firms and a military that fostered new management techniques like P.P.B.S (Planning, Programming, Budgeting Systems), which was installed first in the Pentagon after McNamara left Ford Motors to become secretary of defense in 1961.

These are the negatives; among the positives is the firm-centered discipline BWL inherited from the past. Managers in German firms, struggling to reestablish themselves, were suspicious about the usefulness of a BWL professorate that tried to make management a science. The issue was not BWL as a *Kunstlehre* but BWL as *Wissenschaft*. The encounter in 1965 between the businessman Hans Dichgans and a group of twenty BWL professors to thrash out their differences reveals the serious nature of their relational problem. Dichgans complained that academic business economics was no longer a *Kunstlehre* willing to be judged on how well it accommodated business. “Everything must the more so serve an abstract idea of higher science. Our young professors have less and less contact with praxis” (quoted in BWL, 1965, p. 1427, my translation). He recommended that more professors work for extended periods in business and industry and that their students prepare for business rather than academic careers. Students “should learn enough in a three-year course to qualify for a beginning position in the economy, from which they could improve their knowledge through post-experience education” (*Ibid*.).

The professors were not indifferent to the call for more contact between BWL and praxis, but they could not imagine how it could be done as Dichgans outlined. To the suggestion that professors acquire more practical experience, Professor Wolfgang Kilger of Saarbrücken University replied that for both academics and practicing managers the crucial years in a career occurred between ages thirty and forty. Since it took a student four years to become a *Dipl.-Kaufmann*, two to complete a doctorate, and three to write a *Habilitationsschrift*, the qualification needed to be appointed to a university chair, it was not possible for the prospective professor to complete his/her education before age thirty. As a young professor, he/she then had to devote effort to academic science to advance discipline and career – so there was not much time to work in business. The professors also objected to Dichgans’ suggestion that the *Habilitationsschrift* be eliminated as a requirement for a professorship (as engineers did in their *technische Hochschulen*), in order to attract more people from business into academia. One stated “a person who had worked years in accounting was ill-suited to teach academic accounting”. Another pointed out “business does not understand reality, which is getting more complicated every day.” It stands to reason, another professor intoned, that *Praktiker* needed the professors and their science because of its abstractions, or, as still another professor explained, precisely because the scientists are interested in generalities, not like the *Praktiker*, in particulars. Schmalenbach and his followers, with their focus on BWL
as a *Kunstlehre*, would never have framed the debate in such terms, but under Gutenberg’s influence they now did.

Those imbued with the importance of nomothetic science to business in venues like *Universitätsseminar der Wirtschaft* (University Seminar on the Economy), the Koblenz Corporate School of Business, the Otto Bisheim Business School, and certain faculties involved in the construction of market and finance architectures, actively persisted in their advocacy.

But a different outcome from the one they sought for BWL prevailed. People in praxis in league with BWL academics devised an education system that let academics keep control of their bailiwick and *practicing managers keep the theory boys out of their hair*. The system divided the education process in terms of its ability to improve a student’s capacity to learn, (his or her *Fahigkeit*), which the academics assume through theoretical training and a schooling of the mind (*Denkschulung*), and the training needed after business school graduates join a firm to make and keep them ready for the job (their *Fertigkeit*), jurisdiction over which people in praxis assume.

Undergraduate students concentrate on a functional specialty (accounting, banking, marketing, etc.), not management. After initial employment in lower level positions within a firm, the graduate Dipl.-Kaufmann is not put on a fast-track management career to the top like MBAs recruited from elite business schools in the U.S. Advancement in a German firm-centered world does not depend much on the degree people earned or the place where they got it, what counts more is their knowledge of their specialties (*Fachkentnisse*) on the job and their performance (*Leistung*) BWL graduates can end up in top management, based on these criteria, but so can graduates from sub university schools in which the *Dipl-Handelslehrer* teach, or people who have only completed apprenticeship training.

The center of gravity for post-experience education was not in academia. Top management, not the faculties of business economics, organized the *Baden Baden Gespräch*, in which participants learned about general management from practicing managers. Although professors might be asked to participate in the gatherings, promising younger managers were learning about how to manage from their successful superiors; professional associations like engineering societies sponsored short term courses on various subjects, which BWL graduates might attend or in which they might teach if invited. Business and industry sponsored the creation of post-experience nonacademic teaching networks, like that of the Wuppertaler Kreis, which thrived after the war. Dr. Wipperman, business agent for the Wuppertaler Kreis, observed that he or members of his staff do not work with, or show much enthusiasm for working with, faculties of business economics in post experience management education (interview, Cologne, July 18, 1984). To them the academic viewpoint was too theoretical and the professors too divorced from praxis.

The ability of the people in praxis to control post-experience education conjures up Schmalenbach’s definition of business economics as a *Kunstlehre*, a discipline that is tied to something outside itself, in this case not to nomothetic science that is reminiscent of the post-reform regime of business school education in America but to a firm-centric education that shuns a *Lehre von Management*. 

128
The dysfunctionalism of U.S. (B) management education model compared to the functional success of the German (A) model post WWII

This discussion covers educational, economic, and social examples of each model’s relative prowess in its own area. In model (B) education, two particular events upset the Anglo-American greater academic community (a) the failure of nomothetic economic and business studies to become prescriptive sciences and (b) the consequences, compared to those in model (A) with its different academic traditions, of the elimination of the ideographic tradition in economic studies for model (B) business studies. A third (c) is the reaction of people in both models to the crises in manufacturing in the fourth quarter of the 20th century, which greatly affected outcomes in industrial reform.

I discussed (a) the failure of nomothetic science to succeed as prescriptive science as it pertained to operations research in chapter 2, “The New Paradigm Revisited,” Locke (1989). Khurana, after thoroughly describing the implementation of the nomothetic New Paradigms in U.S. business schools, ignores its prescriptive failure, but the pages of the Real-World Economics Review and the journals of the World Economics Association are filled with articles about it. Practicing managers, who live in a world of bounded rationality, incomplete information, and management unknowns, realized the futility of establishing economics and business studies on nomothetic science.

The problem, however, is not the failure of economic departments and business schools to create a prescriptive science, but the refusal of nomothetic neoclassical economists and mathematical modelers in them to admit the failure, and their actions after they gained a monopoly of the sinews of institutional power, that produced dysfunctionality in Anglo American higher education. That dysfunctionalism is expressed in their constant battle with people in academia who realize the prescriptive failure of the nomothetic science project, with which readers of the Real-World Economics Review blog are painfully familiar, and a dysfunctionalism that results in education from their narrow minded refusal (b) to accept the importance of the ideographic tradition in economic and business studies during the current crisis in U.S. management capitalism.

The elimination of ideographic economics from the economists’ ranks is particularly poignant, inasmuch as the nomothetic neoclassical economists’ attitude towards ideographic economics seriously compromises any attempt of economists to evaluate capitalism’s shortcomings. This includes Veblen’s view that modern capitalism produced a set of socially beneficial tendencies but also a set of parasitical forces. Among the beneficial tendencies he counted “workmanship, industry, the machine process, and technological progress” (Schatzberg, 2006). On the parasitic side, he listed “predation, business enterprise, absentee ownership, and other pecuniary institutions” (ibid., 499) For business schools, Veblen’s stress on the parasitical side of pecuniary institutions is especially significant, since the financilization of the economy has been carried out by nomothetic thinking finance professor in business schools in league with private financial and banking firms governed through director primacy, Veblen’s parasitical forces. In The Theory of the Leisure Class: an Economic Study of Institutions (1899), Veblen also laid out a social critique of conspicuous consumption as a function of social class and of consumerism, derived from the social stratification of wealth. Schatzberg observed that the institutionalist Seligman, who defined economics as the study “of the social conditions necessary for the sustenance of life,” (Schatzberg, p. 498) opened therewith the discipline to the ethics of social criticism. The disappearance of
historical and institutional economists ended this sort of analysis among economists of the U.S. capitalist culture.

More importantly, the recent concentration on finance and investor capitalism in business schools at the expense of manufacturing eliminated the possibility of business schools playing any serious role in solving the manufacturing crisis. They actually opted out of the effort. Robert S. Kaplan, former dean of Carnegie-Mellon Business School and then a Harvard Business School professor, after reviewing articles published in leading operations management journals and examining research and teaching in top business schools, found that only one to two percent of the U.S. business schools had “truly been affected, as of early 1991, by the Total Quality Management revolution that had been creating radical change in many U.S. and worldwide businesses,” and was integral to the Japan Production System (Kaplan, 1991, p. 1; Ishikawa, 1985). He concluded that “American business school research and teaching contributed almost nothing to the most significant development in the business world over the past half century – the quality revolution.” U.S. MBA education proved to be at best neutral in that the neoclassical economic theory and prescriptive sciences it devised and taught had very little to do with a people-oriented management processes American production engineers outside business schools sought to introduce into manufacturing.

In fact, the financial reporting systems created in business schools and the people that ran them in praxis frustrated any reform of the production process. H. Thomas Johnson, after evaluating U.S. automobile production management, noted:

Successful [U.S.] managers believed they could make decisions without knowing the company’s products, technologies, or customers. They had only to understand the intricacies of financial reporting ... [B]y the 1970s managers came primarily from the ranks of accountants and controllers, rather than from the ranks of engineers, designers, and marketers. [This new managerial class] moved frequently among companies without regard to the industry or markets they served ... A synergistic relationship developed between the management accounting taught to MBAs in business schools and the practices emanating from corporate controllers’ offices, imparted to management accounting a life of its own and shaped the way managers ran businesses (Johnson and Bröms, 2000, 57).

At first the abstract information compiled and transmitted by these computer systems merely supplemented the perspectives of managers who were already familiar with concrete details of the operations they managed, no matter how complicated and confused those operations became.

Such individuals, prevalent in top management ranks before 1970, had a clear sense of the difference between “the map” created by abstract computer calculations and “the territory” that people inhabited in the workplace. Increasingly after 1970, however, managers lacking in shop floor experience or in engineering training, often trained in graduate business schools, came to dominate American and European manufacturing establishments. In their hands the “map was the territory.” In other words, they considered reality to be the abstract quantitative models, the management accounting reports, and the computer scheduling algorithms (Johnson and Bröms, 2000, 23).
In Germany the monopolistic hegemony that nomothetic neoclassical economics and mathematical modeling gained in U.S. business schools could not occur. The Germans never bought into nomothetic neoclassical economics to the same extent as Americans, for an interest in ideographic-institutional economics, since it is their tradition, has continued, especially in universities located in the former German Democratic Republic, which through their isolation escaped Americanization during the Cold War. Witness the dissertation on Friedrich List that Arno Mong Daastøl successfully defended at the University of Erfurt in 2011. It is a thorough work that not only comprehensively covers List but traces the presence of Listians within a community of ideographic economists into the 21st century.

Nor did business economists housed in the German (A) model of management education try like U.S. business schools to ignore, or worse, if inadvertently, to play a spoiler role when German manufacturing like American faced an existential threat. On the contrary, the intense effort in the first half of the 20th century that BWL professors made to integrate business with technical studies in their degree programs (especially the Dipl.-Wirt-Ing. degree program in technical universities), and the work they had done with engineers in the Frankfort Society and the RKW, promoted an active participation in these late 20th century reform efforts.

BWL Professor Horst Wildemann, teaching courses primarily to engineering students on work-process innovation in the Munich Technical University, led a substantial group of over 100 research consultants (30% with BWL degrees, 50% with Dipl. Wirt-Ing. degrees, 20% Dipl.-Ing.), which included 35 graduate assistants. Their work was heavily oriented to mathematical modeling and computer simulations; their task to help German firms meet the Japanese challenge by adopting the Japanese Kata-like production culture and its techniques. By 1994 Wildemann’s team had already introduced Japanese production processes in 200 European (mostly German) firms, including Daimler-Benz, Grundig, Philips, and Volkswagen. At Volkswagen, his group spent three years teaching small-group quality control management techniques in five-day courses to over 2,500 managers (Locke, 1996, pp. 199-201). Through them 30 to 50 percent of German industry had by 1994 successfully implemented Total Quality Management, including Just in Time, Kaizen, and/or other Japanese work-process techniques. Notice the educational composition of the work teams. BWL through its Diplom Wirtschafts-Ingenieur engineering education tradition made a significant contribution to what turned out to be a successful German response to the Japanese challenge (interview with Professor Dr. Horst Wildemann, Munich, 24 July 1994).

The success of this German effort to save their manufacturing industry, in which BWL educational tradition embedded in the (A) model expressed itself, and the relative failure of the Americans, in which the nomothetic neoclassical economic led reforms of elite U.S. business school expressed itself, can be discerned through comparative analyses of the top twenty firms in each country, ranked by revenues in 2012.
In the U.S. these firms were:

1. Exxon
2. Wal-Mart
3. Chevron
4. ConocoPhillips
5. General Motors
6. General Electric
7. Berkshire-Hathaway
8. Apple
9. Ford
10. Hewlett-Packard
11. AT&T
12. Valero Energy
13. Bank of America Corp
14. McKesson
15. Verizon Communications
16. JP Morgan Chase & Co
17. Fannie Mae
18. CUS Caremark
19. IBM
20. Citi Group

(Source: Stahl, 2013, 59)

In Germany the top twenty firms were:

1. Volkswagen
2. E.ON
3. Daimler
4. Siemens
5. BASF
6. BMW
7. Metro
8. Schwarz
9. Deutsche Telekom
10. Deutsche Post
11. Aldi Group
12. BP Europa SE
13. Robert Bosch
14. RWE
15. Rewe Group
16. Edeka Group
17. Audi
18. Thyssen Krupp
19. Deutsche Bahn
20. Bayer

( Ibid., 61)

Some firms on each list are classifiable under the same rubric, e.g., retail giants (in the U.S., Wal-Mart and McKesson; in Germany, the Aldi and Edeka Groups). Others are famous oil and energy firms, mostly on the U.S. list. But there are two big differences between the lists that are of interest here. One is that among the top twenty U.S. firms there are many drivers of financialization (Berkshire-Hathaway, Fannie Mae, Bank of America, JP Morgan Chase Co, Citi-Group, and GE Financial), or U.S. firms that are the creation of financialization (Hewlett-Packard: IPO 1957; Apple: IPO 1980). On the German list, there are none, i.e., not one is a financial institution, not one is a stock market IPO creation.

The second significant difference gleaned from a comparison of the top twenty German and U.S. firms pertains to manufacturing. Few of the manufacturing firms on the U.S. list were famous before World War II (Ford, GM, GE), but such firms dominate the list of the German top twenty, many of them prominent even before World War I (Deutsche Post, Robert Bosch, Daimler, BASF, Thyssen Krupp, Bayer, and Deutsche Bahn). Whereas the efforts of business economists to save firms that had been resurrected after WWII succeeded, the old stable industries that dominated the U.S. economy in 1960 had disappeared by the 2000s, thanks in large part to the role business schools played in financialization. Was this dysfunctionality? Considering the misdistribution of wealth that has occurred in the U.S., the nomothetic thought patterns, derived from the operations research methodologies adopted by neoclassical economists and employed by firms in financial market modeling, are a disastrous consequence of their victory over the ideographic economists post World War II.

A final point is about the catastrophic effect that the elimination of ideographic thinking in economics had on firm governance, again especially during our era of financialization. Americans have a proprietary conception of the firm and a simple idea of its purpose,
meaning that the firm is a money mill in which success is measured by return on investment and managers are agents of the capitalist investors, deemed efficient when they maximize profits and stock market valuations climb. The nomothetic neoclassical economics taught in business schools spawned a form of management thought that ignored the fate of other stakeholders in a firm, especially in times of economic crisis. Hedge fund operators and buyout firms specialized in the elimination of legacy costs, like union retirement plans and medical benefits, which wiped out defined revenue pension plans in private U.S. firms between 1980 and 2000; business school graduates devised derivative packages and mathematical financial marketing schemes that, in the subprime mortgage housing crisis, ushered millions into bankruptcy. Nomothetic neoclassical economics proved inadequate to analyzing the shortcomings of the instruments it had employed in the economy and society, and employees had no power to defend themselves.

The social dimension of the BWL educational model projected into the German Federal Republic

For the first generation of German BWL professors, social issues came up in their discussions about firm efficiency (Wirtschaftlichkeit). For Americans it has to do with a firm’s Rentabilität as a money mill (Geldfabrik), which was easy for the accountants to measure. Wilhelm Rieger at the Nuremberg HHS and his followers accepted that. Professor Heinrich Nicklisch, in HHS Berlin, demurred. If he accepted the efficiency principle, the efficiency yardstick was not the greatest income to the capitalist at the least expense, but the greatest benefits to the community. Nicklisch, one of his students wrote

stressed the creative character of human work, refused, correspondingly to accept the liberal wage system as a proper basis for wage payment, demanded profit-sharing for workers and employees, fought the idea that profits were justified when the capitalists did not work, and favored the workers’ participation in management expressed in the regulations and business committee laws of 1920 and article 165 of the Weimar Constitution which set up workers and economic councils at the firm, district and national level of the economy (Nicklisch, 1921).

Nicklisch’s reference to workers’ and employees’ participation is the key point, for even if BWL professors agreed that efficiency had to be more than the greatest return to the investors, as Schmalenbach and others did, thereby disagreeing with Rieger, they did not know how, technically, to measure firm efficiency in terms of satisfying community instead of just investors’ interests. Therefore, many fell back on a stakeholder concept of the firm, in which employee and worker interests would be legally spelled out and defended by giving them a voice in firm management.

Emperor Wilhelm II had incorporated the concept into a speech in 1890, asking for the creation of worker-representative bodies within factories that would defend employee interests in negotiations with employers. The resultant Law for the Protection of Labor, which the reformist Friedrich Lange sponsored in the Reichstag, granted workers joint consultation rights (Mitberatungsrecht) on social matters. This was not co-determination (Mitbestimmung), but the law authorized the organization of plant committees in all factories covered by the Industrial Code of 1869 if they had more than 20 employees (Schuchman, 1957, 14).
Employers were reluctant to accept co-determination, which the people who drafted the Weimar Constitution tried, as Nicklisch mentioned, to enshrine in the basic law. Since Hitler did not embrace it in his brutal regime, Germany had to await the restoration of national sovereignty in the Bundesrepublik for the co-determination laws (1950-51, extended by legislation in the 1960s) to be passed and institutionalized.

American management schools took up human rights issues in the name of efficiency, exemplified by Harvard Business School’s Elton Mayo studies of the Western Electric Company’s Hawthorne Works, in Cicero, Illinois, that began in 1924. But the business schools’ human rights investigators, including those specifically involved in the Hawthorne studies, never thought human rights in firms included the participation of employees in firm governance; on the contrary, the human rights movement in U.S. business school education was anti-union, anti-employment participation throughout. It was interpreted as an expression of management enlightenment evoked by the Harvard Business School investigations. (See Ch. 6 the “Discovery of the Human Worker,” in Cumming et al, A New History of Management.)

In Germany, the U.S. general in control during the occupation, Lucius Clay, did everything to forestall the passage of co-determination legislation. After the laws passed, American businessmen attacked them continually. During a visit to the Ford plant in Cologne, after co-determination was legislatively extended in 1976, Henry Ford III stated that the law violated management’s right to manage.

The co-determination laws, because they affected almost everybody in the work world, required a broad spectrum educational effort about them in order to be operative. The most important of union-involved education in this respect was undoubtedly IG Metall’s education center at Sprochhovel, opened in 1971. In the first two years of operation, Sprochhovel welcomed 25,000 visitors into its seminars, among them 2000 works councilors and youth representatives in seminars just on labor’s legal rights under co-determination. (See, Locke, 1996, Chapter on German obstinacy, which describes the “Emergence of the German Management Alternative.”) Professors and students of BWL participated; they designed special courses on co-determination for their students; they taught post-experience courses about it to managers, works councilors, and members of supervisory boards in nonacademic venues. Did they succeed in helping to create a different German management culture from the American? BWL Professor Alfred Kieser, of Mannheim University, thinks they did, and he told a story when I interviewed him in 1994 that illustrates how much.

At a restaurant in Rome, Kieser was disturbed by a raucous party in a nearby room, fellows of the rough and tumble type. Inquiring about them, he was told that they were a delegation of the American Federation of Labor. Kieser, when telling the story, sought less to disparage the Americans than to emphasize how in a room of German managers and works councilors it would be impossible today to tell the managers from the employee representatives (the works councilors); both sides wear the same clothes, carry the same briefcases, speak the same language, and in many cases have the same education. Kieser’s Ideal-type picture of an employee-manager setting, is a far cry from the ideal-type of U.S. union leaders portrayed in films like On the Waterfront (Professor Dr Alfred Kieser, interview in Mannheim BWL faculty, 14 July 1994).

But did co-determination succeed as management? U.S. managers and business school professors generally do not believe employee participation can produce “efficient”
management. BWL Professor Horst Wildemann, involved in the transformation of German manufacturing in the 1980s and 1990s, disagrees. In four years at Volkswagen he worked closely with works councils and IG Metall shop stewards. The works councilors in his words were "very intelligent people," who fully appreciated the need to improve work processes, but also understood the impact that the changes would have on jobs numbers in the workplace and on the need to reduce work time and pay. He noted that his group taught the new techniques to the shop stewards at the same time that they taught them to management, and that the union (IG Metall) not only promoted the implementation of Just-In-Time and other work processes but often led management instead of following it in their adoption (interview with Professor Dr Horst Wildemann, in his Munich home, 21 July 1994).

If the BWL tradition as a Kunstlehre has succeeded so well in promoting good firm governance, within the framework of co-determination, the triumph of nomothetic science in U.S. business schools, under director primacy firm governance, has not.

Conclusion

So what, as Spender suggests, can "folks" involved in the Anglo-American (B) model of management and business school education learn from the older German (A) model? They cannot learn from model (A) that the nomothetic mathematical models have spectacularly failed. Since the housing and financial crises of 2007-2008 everybody knows that, even some of the chastened modelers. But, they can learn, against a background of the cooperative failure of elite U.S. business schools to participate vigorously and fully in the renewal of manufacturing in the 1980s, and the failure of business school conceived nomothetic investor models of financial markets to which business school energies had turned, that German business economists, in a tradition of collaboration with industrialists and engineers, made a significant contribution in praxis to sustaining the wherewithal, and even reshaping it, of German industry.

They can also learn how important systems of business studies can be in determining economic, social, and political outcomes.

In the U.S. MBAs from the top business schools are part of a privileged order increasingly working in the finance and investor sector of the economy or working for it; they have joined the ranks of the top 10 percent of income earners, who are responsible for the gap increasing between the rich and the poor. The prosperity of the MBA elite does not depend so much on their knowledge as the power to decide who sets wages, salaries, and bonuses. That power resides in the shareholders, to some extent, but principally in director primacy management that determines who gets what, and the bulk of the what goes to management. Two features of the system that is immersed in stock market casino capitalism: (1) The business schools and their graduates do not challenge but promote it. (2) Those who work in firms without a voice in decision making, are victims of it.

Director-primacy firm governance produces the shocking differentials in pay between top managers and non-management employees in U.S. firms, the greatest in modern economies, and policies that reward shareholder at the expense of employees like stock buyback schemes. The wealth disparity gap they promote is responsible for dissolving the social cohesion essential to the survival of American democracy.
Susan R Holmberg, director of research at the Roosevelt Institute, in an article written with Mark Schmitt, “The Milton Friedman Doctrine is Wrong: Here’s How to Rethink the Corporation,” stated that Americans will not be able to fix the problem of the increasing gap between the incomes of the very rich and the middle classes until they address the nature of the corporation. Holmberg and Schmitt suggest Americans adopt the German system of co-determination.

Why the business schools did not fix the problem with co-determination has just been explained, why the Germans did can be gleaned from a good look at the (A) model. BWL exists in an inclusive culture of Technik, not one of financialized elitism. The presence of BWL graduates in many echelons of a firm who are well-schooled in co-determination rights permits them to participate on both sides of governance, as employee representatives on works councils or as representatives of management in order to achieve good results for an equitable distribution of the rewards of enterprise to all firm stakeholders, and to spare their firms and country the worst inequities of director-primacy MBA capitalism.

Bibliography


*Zeitschrift für handelswissenschaftliche Forschung*, founded in 1906, continued after WWII as *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung.* Published by the Schmalenbach-Gesellschaft--Deutsche Gesellschaft für Betriebswirtschaft; Schmalenbach-Gesellschaft zur Förderung der Betriebswirtschaftlichen Forschung und Praxis.

Author contact: blocke@hawaii.edu, lockehotel@gmail.com


You may post and read comments on this paper at [https://rwer.wordpress.com/comments-on-rwer-issue-no-83/](https://rwer.wordpress.com/comments-on-rwer-issue-no-83/)

137
Does the maximization principle break down during recessions?
Philip George [India]

Abstract
This paper shows that the principle of profit maximization by firms does not hold during recessions. It also offers evidence to show that the principle of utility maximization by consumers collapses during recessions.

JEL Classifications D11, D21, E12, E13

Keyword: profit maximization, utility maximization, Keynesian, asset markets, recessions

Introduction
The principle of maximization is central to economics. It is what makes economics mathematically tractable. Paul Krugman (Krugman, 1996) speaks for most mainstream economists when he says:

“For there is no question that conventional economics has gone beyond the general ideas of intelligence and interaction to a much harder-edged, extreme formulation. At least since Paul Samuelson published Foundations of Economic Analysis in 1947, the overwhelming thrust of conventional theory has been to say that agents are not only intelligent, they maximize — that is, they choose the best of all feasible alternatives. And when they interact, we assume that what they do is achieve an equilibrium, in which each individual is doing the best he can given what all the others are doing.”

And later on, after agreeing that the assumptions of maximization are not often very realistic, Krugman says:

“Personally, I consider myself a proud neoclassicist. By this I clearly don't mean that I believe in perfect competition all the way. What I mean is that I prefer, when I can, to make sense of the world using models in which individuals maximize and the interaction of these individuals can be summarized by some concept of equilibrium. The reason I like that kind of model is not that I believe it to be literally true, but that I am intensely aware of the power of maximization-and-equilibrium to organize one's thinking — and I have seen the propensity of those who try to do economics without those organizing devices to produce sheer nonsense when they imagine they are freeing themselves from some confining orthodoxy.”

In other words, maximization is one of the two principles that set apart true economists from mere pretenders.

In what follows I consider the parallel situations of profit maximization by firms and utility maximization by consumers, and show that both break down during recessions.
A. Profit maximization by firms

The figure above depicts the argument for profit maximization by firms under perfect competition. Assume that the firm is currently at point A. Since the marginal revenue (MR) is higher than the marginal cost (MC), the firm can increase its output and its profit to the point B where the marginal cost curve intersects the marginal revenue line. Beyond this point the marginal cost is higher than the marginal revenue and increasing output would reduce its total profit. So the firm should increase its output up to B which is the point of profit maximization.

This argument assumes that there are no barriers to increasing output from point A to point B. However, by definition, recessions are periods in which demand is falling. So the typical firm that attempted to increase its output from A to B during a recession would find the additional output remaining unsold. If it tried to attain the point where theoretically its profit is greatest it would instead find its profit falling drastically. The argument for profit maximization implicitly assumes that output is always at a point where it can be freely increased. This assumption is invalid during a recession and hence the principle of profit maximization collapses during recessions.

This is irrespective of other possible lines of critique, for example matters of actual firm practice and the construction of curves (early Sraffa) or the structure and strategies within markets based on imperfect competition (Joan Robinson and so forth), or the varieties of alternative accounts of capitalist firms and strategy (in a spectrum spanning Baran and Sweezy, Dunning, Hymer, Vernon and others).

It is interesting to look at how General Equilibrium theory deals with the idea of profit maximization. Theory of Value (Debreu, 1959), one of the classics of GE theory, simply states that firms act so as to maximize profits. But Kenneth Arrow (Arrow, 1972) in his Nobel Prize
speech, in arguing that part of a particular equation signals profit maximization, says that the assumption is that production is independent of consumption (George, 2016). This is exactly the condition depicted in Fig 1. Profits can be maximized only if it is assumed that consumption does not place any restrictions on expanding production. If enhanced production cannot be consumed the principle of profit maximization fails.

Early in his Nobel lecture, Arrow asks rhetorically: “The fundamental question remains, how does an overall total quantity, say demand, as in the Keynesian model, get transformed into a set of signals and incentives for individual sellers?” To every capitalist who experiences a recession the answer is clear. When he finds that part of his output remains unsold and that lowering prices does not help to clear his unsold inventory, he realizes that the market is signaling him to reduce output. In asking his question, Arrow’s assumption is that production is independent of consumption, an assumption that he states explicitly later in the speech.

To understand why increased production is not consumed one must turn to the other half of the maximization principle, utility maximization.

B. Utility maximization by consumers

The figure above shows utility maximization by an individual who consumes only two goods, A and B. The line segment MN is his budget line. At M he spends his entire budget on good B. At N he spends his entire budget on good A. At intermediate points he buys some of both goods.

The green lines are his indifference curves, which are curves of constant utility. Points above the budget line are financially unachievable. Points below the budget line are sub-optimal.
Utility is maximized at point Q where an indifference curve is tangent to the budget line. Next we examine what happens when the budget line moves.

In the above figure the budget line has fallen from its initial position MN to CD. (Prices have remained unchanged which is why CD is parallel to MN.) In response the consumer adjusts his consumption of the two goods so that his new indifference curve is tangent to the new budget line at R. If the budget (income) decreases and the consumer responds by cutting his consumption so as to be tangent to the new budget line the principle of utility maximization is not negated. However, if the consumer reduces his consumption even when his budget (income) has not shrunk we can conclude that he is not maximizing utility.

There are thus two possibilities. The first is that income falls and the consumer reduces his consumption accordingly. In this case utility maximization continues to hold. In the second situation the consumer reduces his consumption even when his income has not fallen. In this case the consumer is obviously not maximizing utility.

When can these two situations arise?

There are two possible models. The first is the Keynesian model. For some reason, the animal spirits of capitalists get deranged and they reduce their investment. This reduces the income paid out to workers who reduce their consumption accordingly.

The second is the asset price collapse model. A collapse of one or more asset markets (e.g. the housing and equity markets) knocks a big hole in the accumulated saving of consumers who then increase their current saving (reduce consumption) in order to rebuild their lost accumulated saving. This happens even when their income has not fallen.
In many conventional new classical models agents have infinitely long lives. In reality agents have finite lives and finite working lives. During their working life they save money in order to prepare for large planned and unplanned expenses and, most importantly, to provide for that part of their lives (retirement) during which they will not receive wages. Agents invest their savings in various financial assets, including housing. When asset markets rise, the value of their accumulated saving (after considering increases in asset prices) increases, and they may feel confident enough to reduce their current saving. When asset markets collapse, their accumulated saving (net worth) takes a big hit. There is only one way they can repair the damage to their accumulated saving and that is by increasing current saving (or reducing consumption). They cannot opt for credit because that would further reduce their net worth; credit constraint is not the problem. Accumulated saving, net worth and wealth are thus synonyms in the above model.

Which of these two models is the correct one? Or to put it another way, is the fall in consumption caused by a fall in income (which in turn is caused by a fall in investment by firms) or is the fall in aggregate income caused by a fall in consumption by individuals? In the rest of this paper I examine the evidence on both sides.

C. The evidence

When a sizeable number of individuals respond to a collapse in their accumulated saving by reducing their consumption the effect on the macro economy is a contraction in aggregate demand. In response capitalists reduce both output and investment which in turn results in a cut in factor payments which further reduces consumption. And so on.

What comes first, a cut in aggregate income or a cut in aggregate consumption? Unfortunately, we cannot just measure these two variables to arrive at an answer because one person’s cut in consumption means an immediate fall in some other person’s income. The two occur simultaneously. We must look elsewhere for evidence.

1. The ineffectiveness of near-zero interest rates

For years after the financial crisis of 2007-09 the Federal Reserve Board of the US (and the central banks of other countries) held the Federal Funds Rate (and related rates in other countries) close to zero. But the economic response has been far less positive than expected and economies have not returned to their former growth path.

In theory, since recessions are caused by a fall in investment, pushing interest rates down should make investment in otherwise unviable projects viable. But cutting interest rates hasn’t helped much in reality. Nor have huge injections of cash through the medium of asset purchases by the central bank. Keynesians usually explain this by drawing on the liquidity trap, although Keynes himself expressed doubt about whether it ever comes into play.

But the simplest explanation for the lack of response of investment to near-zero interest rates is that the recession was caused not by a fall in investment but by a fall in consumption, and the latter is not very responsive to interest rates.
2. The long-drawn-out recoveries that follow asset market collapses

In textbook presentations of macroeconomics \( Y = C + I + G + X \), where \( Y \) is the GDP and \( C, I, G \) and \( X \) stand for aggregate consumption, aggregate investment, government expenditure and net exports.

Since a recession, in the Keynesian model, is set off by a drop in investment, all one needs to do is to compensate for the fall in investment by raising government expenditure. Time does not enter into the above equation at all. But the reality is that time is very much a factor in recoveries that follow asset market collapses. These tend to be long-drawn-out, the Great Depression and the Great Recession being classic examples.

For an explanation we need to look at the effect of the asset market collapse on accumulated saving. According to the Federal Reserve Board’s Survey of Consumer Finances (Federal Reserve, 2012) the median family net worth after the crash fell in 2010 to levels not seen since the 1992 survey. The median family thus lost 18 years of accumulated saving. To recover that lost wealth, the median family would have had to double its saving rate for 18 years, all else remaining equal.

In the equation \( Y = C + I + G + X \), a drop in consumption has the same effect as a drop in investment. Mathematically, the two are indistinguishable so far as their effects are concerned. In neither case does time figure as a factor. However, the result of a large and sustained collapse in asset markets is not a one-time drop in consumption but a drop in the rate of consumption as a percentage of income (which is what an increase in the saving rate translates into) for a sustained period. That is where the element of time comes in. That is why recoveries following asset market collapses are long drawn-out.

The Volcker-induced recession too was severe but it was caused by firms not having money to produce goods and services. This in turn affected the incomes paid out to worker-consumers and thus their consumption. But since it was not accompanied by a collapse in asset markets consumers did not have to reduce their rate of consumption for an extended period. Once the Fed loosened its grip on money, firms got access to funds and raised production, so consumers regained their old income and resumed their old rate of consumption.

Fig 4 shows the saving rate for the US from 1959 onwards. From 2.5% just before the start of the 2007-09 recession, it stays at 6% for an extended period thereafter. After the Volcker recessions, in contrast, the saving rate quickly returned to normal because consumers did not need to save more to compensate for losses from a sustained asset market collapse. The shaded columns in the graph indicate periods of recession. The data source is the St Louis Federal Reserve Bank web site.

According to the Federal Reserve Board’s Survey of Consumer Finances (Federal Reserve, 2017), the median family net worth rose 16% between 2013 and 2016. This compares with a decline of 40% between 2007 and 2010 and a decline of 2% between 2010 and 2013. In terms of thousands of 2016 dollars, median net worth in 2016 was 97.3 which compares with 90.6 in 1995 and 105.8 in 1998. Therefore the median family net worth is still roughly what it was 20 years ago. The past ten years have made no difference at all to the real net worth of half the US population. We may therefore expect the consumption rate (as a proportion of income) of the median US family to remain depressed for a very long time to come, with a
concomitant effect on the growth rate of the US economy. The mean net worth, however, has recovered and is now at its highest ever. The contrast between the median and mean measures is of course a reflection of growing inequality. Whether increased consumption by the wealthy will compensate for the depressed consumption of the rest of the population remains to be seen.

In the Great Depression, the immediate effect of the stock market crash was a steep fall in the consumption of consumer durables (Romer, 1990). The reason for the initial fall in consumption being restricted to consumer durables probably lay in the structure of stock ownership at the time of the crash; although 74% of stocks were owned by households (Green, 1971), fewer than 2% of households owned stock (Romer, 1990). By the end of 1933 stock ownership by households had fallen to 56%.

The fall in consumption was followed by a precipitous fall in investment with consequent effects on aggregate demand. But a full understanding of the depression requires a look at monetary factors, especially the large scale bank failures. To an individual, it would not matter whether he lost a substantial part of his saving in the stock market crash or in a bank failure. The effect on his consumption would be the same.

3. The ineffectiveness of fiscal stimulus

Between 2005 and now, the cumulative fall in private residential fixed investment in the US (the putative cause of the recession) has been less than $4 billion. During the same period, the cumulative increase in the budget deficit compared with 2005 is more than $4 billion. The fall in residential fixed investment has been more than compensated by additional government expenditure. But in addition there is also supposed to be a multiplier in operation, though there is some dispute about its value.

Mathematically, the value of the Keynesian multiplier is $1/(1-\text{MPC})$, where MPC is the marginal propensity to consume. The MPC is the ratio of the change in consumption to the change in income for a given period. During a recession, aggregate income falls, by definition. At the same time the saving rate increases, as seen in Fig 4. Thus, the fall in consumption is...
greater than the fall in income. Therefore the MPC is greater than 1. Hence the Keynesian multiplier during a recession is negative, which is not at all what conventional wisdom holds.

A negative value of the Keynesian multiplier does not mean that fiscal stimulus is useless. Rather, it means that the fall in accumulated saving as a result of the asset market collapse is so large that any additional government spending is, in aggregate, completely saved. Even after the recession, the Keynesian multiplier is lower than it was before the recession.

D. Conclusion

It is easy to show that the principle of profit maximization by firms cannot hold during recessions. Purely from considerations of symmetry, one would guess that the principle of utility maximization too cannot be valid during recessions. And as we have shown above, there is considerable macroeconomic evidence to conclude that this is the case.

But even when there is no recession there is reason to doubt the principle of maximization. We argued that if the budget line falls and the consumer reduces his consumption as a consequence, the principle of utility maximization holds. But as we have seen, what is commonly taken to be an objective budget line is actually a subjective spending line. The consumer first decides how much to spend from extraneous considerations (such as the value of his accumulated saving) and then decides how to derive the best results from the subjective decision so made. So even during non-recessionary periods the principle of utility maximization can be questioned.

Similarly there is no reason to believe that every firm during non-recessionary periods is free to expand its output right up to the point where marginal cost equals marginal revenue.

The principle of maximization is best thought of as a useful mathematical device when the economy is growing at a rapid clip.

We may mention in passing that if the principle of utility maximization does not hold, the Euler equation, which is one of the fundamentals of new classical economics, would have to be abandoned.

References


Author contact: philip.george@gmail.com


You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-83/
Abstract
In the mid-1990s, liberalization of the capital and financial services markets was seen as the flagship of structural reforms within the EU. This remained relatively uncontroversial until the outbreak of the 2008 global financial crisis, which seriously undermined confidence in financial capital and “free markets”, making the European public skeptical of austerity measures based on the liberalization project. Yet the swiftly engineered shift in attention from the abuses of unregulated markets to the sovereign debt crisis changed the agenda, and paved the way for the completion of the original plan, bolstered by anti-inflationary and pro-market measures. Though capable of yielding short-term profits for northern European capital, this approach has seriously undermined the continent’s already weakened social model. In what follows I shall focus on the transition from the aggressive market-oriented integrating strategy inaugurated in Bologna and Lisbon in the late 20th century to the new pro-market policies adopted following the crisis, which have widened and deepened the gap between Eurozone member states. I shall also focus on a less self-evident parameter of recent developments: the prevalence of technocratic solutions, and the retreat of political processes and elective bodies in the formulation and implementation of economic policy in Europe.

1. Introduction: what happened to the Europe of knowledge?

The debate over the growing need for the European Union (EU) to adjust to an increasingly cut-throat, globalizing world is neither original nor new. Discussion on the issue began in the late 1990s, when 29 European Ministers of Education met in Bologna to proclaim the creation of “a Europe of knowledge […] capable of giving its citizens the necessary competences to face the challenges of the new millennium”. The Bologna Declaration emphasized “the importance of education and educational co-operation in the development and strengthening of stable, peaceful and democratic societies”. The explicit aim of making European universities “acquire a world-wide degree of attraction equal to our extraordinary cultural and scientific traditions” reflected a discussion on the contribution of American culture and universities to the USA’s position as the world’s dominant power throughout the second half of the twentieth century. Given the intimate relationship between education and culture, moving faster towards more competitive education systems was thought to be a decisive step towards European preponderance, inasmuch as “the vitality and efficiency of any civilization can be measured by the appeal that its culture has for other countries”. Higher education was destined to be the workshop in which the future world dominance of Europe would be forged.1

Official European texts of the time stressed the belief that raising performance in education would also have a positive impact on “employment, health, social inclusion and active

citizenship”. However, in view of the “strong national nature” of European educational systems, policy intervention should encourage “the comparison of performance in education across Europe” and “greater transparency in the recognition of qualifications and periods of study and training”. The inexorable question was how these targets could be attained against a background of restrictive fiscal policies. Indeed, due to budgetary constraints, over-optimistic forecasts and weak economic growth, it proved impossible to fulfill earlier lofty promises to increase per capita investment in human resources, reduce school failure and drop-out, and raise the share of graduates in mathematics, science and technology. The same was true of the proportional increase of women interested in science and technology disciplines, and higher public investment in education. Public expenditure to GDP ratio in the EU trended upwards from 1970 (30.7%) to 1995 (45.7%). The strengthening of fiscal discipline associated with the Maastricht Treaty (1992) and the Stability and Growth Pact (1997) bear the lion’s share of responsibility for the reversal of this trend during the 2000s. In 2006, Euro area public expenditures were 3.5% of GDP lower than in 1996.

Late 20th century EU documents look on the United States and Japan as Europe’s leading competitors. Spending on research and development (R&D) testified to the widening gap between the EU and other growing industrial economies. Particularly as regards business-funded R&D, the continent’s investment deficit was greater towards Japan, which devoted 3% of its GDP to R&D in 2000, compared to 2.7% for the United States. However, after a long period of economic stagnation in the 1990s and in view of “the problems of the Japanese financial system, which […] obscured the benefits of its high R&D intensity”, Japan lost its role model status in EU texts to China and South Korea.

Funding of higher education institutions in Europe has traditionally relied on the public sector. In 2000, EU public expenditure on education as a proportion of GDP was comparable to that of USA and higher than that of Japan. Available data for 2008-2014 shows an increase in the “innovation divide” within Europe: at one end of the spectrum, public funding for universities rose steadily in countries such as Norway, Sweden, Germany, Austria, Belgium, Denmark and Poland. At the other end, countries in Southern and Eastern Europe (Greece, Hungary and Latvia) implemented swingeing cuts often exceeding 40%. However, one key factor in the post-Bologna reforms was increased private sector involvement in higher education funding. Indeed, in 2004 private investment in educational institutions in the EU was only a quarter of that in the USA (0.6% vs. 2.2% of GDP) and half of that in Japan (1.2%). Increased private investment in R&D, adult education and vocational training was considered crucial to bridging the gap. Universities were called upon to become more sensitive to their varied consumers’ demands, so as to generate income from commercial activities. This trend towards marketization has led to an increase in private expenditure as a proportion of higher education funding, even if against a backdrop of dwindling public funding per student in many European countries.

---


Taking into account the above restrictive framework, there was one area in which intervention by the EU was relatively successful, namely in “the creation of […] a skilled and adaptable labour force on more open and more accessible European labour markets”. This aimed to tackle the high levels of unemployment, poverty and temporary and precarious employment in the EU. Indeed, the issue of joblessness lay at the core of European educational policy texts throughout the 2000s, with particular emphasis being placed on young people, women and the long-term unemployed. The outlook was not bright, however, given the sluggish growth rates in several EU countries. In February 2005 the Council of EU pointed out that “some 33 million people in the EU are not in active employment”, that “Europe’s workforce is ageing, and 32% of the working age population (78 million) is low skilled”, while available forecasts suggested a sharp decline in demand for low skilled workers by 2020 and tougher international competition for highly skilled labour.

In this context it is remarkable that EU texts admitted (albeit indirectly at times) that the de-industrialization of Europe would continue as the 21st century moved forward, and that very little could or should be done at the intergovernmental level to prevent it. What rendered the development of higher education more critical than ever before was the prediction that the Union’s manufacturing base would continue to shrink over the decades to come, and that future growth and social welfare would thus increasingly rely on knowledge-intensive industries and services. Indeed, coupled with an increase in the share of service-related jobs in total employment, the shift in demand from low to high skilled workers rendered general access to European education systems all the more essential in the war on unemployment. Simultaneously, however, increased poverty and inequality in Europe clearly worked in the opposite direction. Even prior to the global financial crisis, political leaders took the view that they should not interfere in the markets to stand up for whatever part of the EU’s manufacturing base had fallen behind in international competition.

According to early 21st century EU reports and texts, Europe should aim to recruit, train, and bring its best brains together, while also fielding leading minds from other regions, in order to make the real differences needed to remain competitive. The continent’s competitive advantage should be based on a well-educated, innovative and trained labor force, rather than on a productive model based on low wages, or on the availability of cheap energy and

---

raw materials, as is the case in many African, Asian and Latin American countries. Yet the years since the onset of the 2008 crisis have seen a gradual watering down of the once firmly held belief that Europe could never hope to win the competitive race against other regions as regards cheap labor, but that it could maintain a competitive edge in knowledge and innovation.

2. East Asian development, and the lessons we have not learned from the past

The debate over the sources of East Asian growth and the measures adopted in response to the “Asian crisis” of 1996-1998 are by no means irrelevant to approaches taken to the current crisis. Some of the fundamental problems inherent in simplistic analysis of such complex phenomena as economic development can be seen in the absence from the so-called European “bailout package” of clauses relating not only to recessionary policies, but also to a development agenda, in the belief that the latter is simply a product of free trade, free-markets and private sector stimulus. Mainstream economists attributed the East Asian economic miracle12 to its opening up to markets, competition and capital, and to the fact that the countries concerned devised and implemented economic policies free of the pressures exerted by politics and business; and those same people then went on to blame the 1996 crisis on East Asian deviation from free market principles. Indeed, the term crony capitalism was extensively used in many simplistic western studies attempting to account for the Southeast Asian crisis of 1996-98.13 One measure of just how shallow analyses can be is the failure to mention the existence in many East Asian countries of non-conventional economic regimes combining atypical ownership rights, protectionist trade policies and intensified market regulation. The absence of fundamental questions about the role of the state and the dissemination of foreign capital and technology is similarly evident in current approaches to the Euro crisis.

The issue of technology transfer from one country to another is not, for example, dependent on the absence of barriers to trade, or on ensuring the most business-friendly regulatory environment. It is noteworthy that when attempting to account for development processes in different parts of the globe, undue emphasis on specific reference values such as the ease of doing business yields at best inconsistent results for the world’s fastest growing economies such as China and India, and also reflects an anti-regulatory bias. In this sense, a former senior OECD officer commented that with regard to ease of doing business rankings, “Basically, countries that have no labor regulation whatsoever get the best marks”.14 Another

11 “Mobilising the brainpower of Europe”, ibid.
12 Indicative of European admiration for these “spectacular” growth rates, which were, for China, “above 7% in 14 consecutive years”, were many communications from the Commission of the European Communities. See for example “Brussels. 11.3.2005, COM (2005) final, “A Community civil aviation policy towards the People's Republic of China - strengthening co-operation and opening markets”, in http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2005:0078:FIN:EN:PDF.
example of the inadequacy of pro-market theories in accounting for economic growth is to be found in the deposition to the U.S. House Science Committee by Robert D. Atkinson, president of a research and educational institute for technological innovation, in December 2012. Discussing the impact of international technology transfer on the USA economy, he focused not only on the fact that Asian nations as a group surpassed the United States in R&D investment for the first time in 2008, but also that U.S. firms are required to share proprietary secrets and open R&D labs with domestic firms in China and other East Asian countries, as a precondition for gaining market access for their products and services. He insisted that China is in a position to implement this effective policy of “forced technology transfer” because it is not a market-oriented liberal democracy, and because its economy is large and powerful enough to resist the pressures of multinational corporations.15

In South Korea, which experienced unprecedented rates of economic growth from the 1960s onwards, boosted by inflows of economic aid, technology and direct investment by the USA and Japan, the state organized firms into giant multisectoral business groups (chaebol) and “through its tight control of central banks and investment decisions […] determined when and where firms would enter new markets”.16 However, the state was active not only in providing various types of economic stimulus [“preferential access to credit, capital, and technology licenses”] and a protected domestic market, but also in imposing the requisite discipline: enterprises whose export capacities were restricted by low productivity levels and could not compete internationally were frozen out of credit markets and faced an unfavorable economic environment.17

Presenting the experience of economic growth in several East Asian countries (South Korea, Taiwan, India, China), Ricardo Hausmann and Dani Rodrik (2003) stressed that the development process did not originate from the supply of a comprehensive technological blueprint with the purpose of constructing turnkey plants identical to those of Great Britain, Germany, or the USA.18 Technology transfers usually fail in the absence of “tacit knowledge”, which should allow for the adaptation of imported technologies to the environment and specific circumstances into which they are introduced. Successful industrialization and the adoption of foreign technology call for the discovery of a fairly restricted high intensive sector in the economy whose effectiveness and suitability for a specific country are “uncertain and unpredictable”. The pattern of specialization characterizing many economies is the outcome of a specific kind of learning which is of value to society but yields modest private profits. When the ease of market entry reduces the benefits of innovation given that business profits from investments lie behind social gains and the motives to discover new products or production processes are lacking, the end result is underinvestment and a flagging economy. Under these circumstances a laissez-faire policy leads to inadequate supply of innovation and governments will have to provide leadership and direction over the private sector so as to “encourage entrepreneurship and […] increase the expected payoff to innovation”.19

18 Ricardo Hausmann, Dani Rodrik, Ibid., p. 624-625.  
19 Ricardo Hausmann, Dani Rodrik, Ibid., p. 605, 607, 629.
The unwillingness to draw the necessary lessons from the 1996 crisis in such a way as to prevent the abuses subsequently seen in both the United States and Europe closely resembles the ostrich approach now being taken towards the Eurozone debt crisis. A good measure of this intransigence and the absence of disinterestedness shown by allegedly politically neutral institutions such as independent central banks was manifest in the testimony given by Alan Greenspan to the US House of Representatives’ Committee on Banking and Financial Services in the aftermath of the Asian crisis. Greenspan, then Chairman of the Federal Reserve and architect of deregulation in the US banking industry, reaffirmed that it would not have been appropriate to apply the 1936 Commodity Exchange Act to financial derivatives “in order to […] deter market manipulation and protecting investors”, because “unlike commodities, whose supply is limited to a particular growing season and finite carryover, the markets for financial instruments […] are deep and, as a consequence, are extremely difficult to manipulate”. Greenspan saw “no reason to question the underlying stability of the OTC [over-the-counter derivatives] markets, or the overall effectiveness of private market discipline”. He was also confident that “these instruments provide extensive protection against undue asset concentration risk” and that they “add significant value to our financial structure, both here in the United States and internationally”.20

Seven years later, before Greenspan lost the epithet “maestro” to the distinctly less flattering yet more appropriate sobriquet “Mr. Bubble”, he declared that in case of a fall in house prices “nationwide banking and widespread securitization of mortgages makes it less likely that financial intermediation would be impaired”, and reassured the public that “the use […] of more sophisticated approaches to measuring and managing risk are key factors underpinning the greater resilience of our largest financial institutions”.21 On October 23 2008, faced by an imminent crash of the financial system, he was ready to admit the collapse of “the whole intellectual edifice” which was behind the “modern risk management paradigm”. However, there was no question of this having been the error of a single man or a single organization; to his mind the fault lay with a whole generation which entrusted “vast risk management and pricing systems […] combining the best insights of mathematicians and finance experts [and] supported by major advances in computer and communications technology”.22 Consequently, the errors of financial institutions and of regulatory authorities were pardonable. There were not the errors of a man or a firm, but of science and humanity themselves. Reasonably enough, such blunders should be forgiven and financial institutions should be supported. But the ordinary people who had placed their trust in financial institutions and regulatory authorities, with devastating results for their employment and property, were guilty as charge. For them, virtue and forgiveness could only be won through austerity, insecurity and privatization.

3. Europe’s new microeconomic agenda

By the late 1990s the initial difficulties inherent in the diffusion and commercialization of new technologies, especially in microelectronics, computing and communication, appeared to have been overcome. Indeed, when in March 2000 the Lisbon European Council declared the move towards a Europe of innovation and knowledge, earlier concerns such as those expressed by Robert M. Solow\(^{23}\) and Paul A. David already seemed irrelevant. Economists took the boom from 2001 onwards as evidence that the global economy had entered into a never-ending era of prosperity.\(^{24}\) European leaders lost no time in grasping the lessons to be drawn: “The sudden arrival and growing importance of information and communication technologies […] call for a radical overhaul of the education system in Europe”.\(^{25}\) At the dawn of the 21st century, the paradigms to be followed in transforming Europe into a dynamic knowledge-based economy originated from the U.S.A. Since the late 1990s America had entered into a period of high growth sustained by information and communication technologies, and an unprecedented expansion of services, banking and internet commerce. “I shall argue that international economic policies have been phenomenally successful to date, and have in many ways changed the contour of the world economy itself, mainly for the better”, declared Anne O. Krueger, First Deputy Managing Director of the IMF, in an April 2006 lecture outlining the evolution of the world economy since the closing decades of the 19th century.\(^{26}\)

Since the onset of the recent crisis, the original optimism of the European politicians who laid down the new economic policy guidelines in Bologna and Lisbon has given way to more modest analyses of EU potentials and dynamics.\(^{27}\) Assessing the Lisbon Strategy in 2010, the European Commission noted that reforms in labour markets “have helped protect jobs in the downturn and stem the rise in unemployment, while the Euro area proved to be an anchor for


\(^{24}\) David compared the slow pace in the transformation of industrial process by electric power between 1890-1930 with the slowdown in productivity growth experienced U.S. economy in the 1970s and 1980s.  


macro-economic stability during the crisis". In making such claims, European officials clearly took an upbeat view of the Euro’s performance and prospects, in line with that offered by Menzie Chinn and Jeffrey Frankel (2008), rather than the distinctly more sombre view taken by David Fields and Matias Vernengo (2011). In a “Commentary” published in International Finance when the crisis in the American financial system was at its apogee, Chinn and Frankel compared the protracted decline of the pound and the emergence of the dollar as the major international currency with the rapid acceptance of the euro as the number two international currency not long after its debut. The writers concluded that despite the “strong inertial bias in favour of using whatever currency has been the international currency in the past”, given the current account deficits of the United States and the size and depth of European financial markets, it was highly plausible that the euro could overtake the dollar as the leading international currency. In contrast, writing at the height of the eurozone debt crisis, Fields and Vernengo thought it “very unlikely that the dollar would be replaced as the key currency by the euro, or any other contender, for a very long period”. According to them, the critical element for monetary hegemony was state power rather than market confidence. The odds of the euro replacing the dollar as the key world currency were thus lengthened by the fact that in Europe there was no equivalent to US treasury bills. Added to this, the ECB appeared hidebound by a purely monetarist approach to central banking, and the EU was clearly unwilling to expand aggregate demand.

Nevertheless, in the aforementioned assessment of the Lisbon strategy, the European Commission was frank enough to criticize the inaction of the EU with regard to “critical elements which played a key role in the origin of the crisis, such as robust supervision and systemic risk in financial markets, speculative bubbles especially in housing markets”. Echoing, at the same time, the usual blame game against the South of the Eurozone, criticism was voiced against “credit-driven consumerism which, in some Member States, combined with wage increases outpacing productivity gains, fueled high current account deficits”. The Commission also expressed dissatisfaction about the suitability of “ambitions endorsed at the highest political level” on the grounds that these “have not always resulted in faster decision making or in avoiding dilution”. This seemingly “innocent” observation would have far reaching effects in the planning and implementation of austerity measures towards EU member states under EC/ECB/IMF supported programs (see below).

Key European Union documents refrain from observing which countries or social groups have benefited the most from Europe’s monetary union, while glossing over the flaws inherent in the design and structure of the single currency. Instead, they insist on the urgency of broadening and deepening macroeconomic surveillance so as to render Europe capable of standing up to international competition. Proposed policy directions include wage restraints, budgetary discipline, deregulation and measures that weaken the economic security and bargaining strength of organized labor.

31 Lisbon Strategy evaluation document, ibid., p. 5.
Christian Thimann’s article “The Microeconomic Dimensions of the Eurozone Crisis and Why European Politics Cannot Solve Them” (2015) is a good example of a text focused on the above objectives.\footnote{32} The fact that Thimann served as a top monetary policy adviser to the ECB president\footnote{33} perhaps accounts for the fact that both Mario Draghi and Jean-Claude Trichet are “privileged commentators” on the article. In approaching the Eurozone crisis, Thimann concentrates on factors not directly associated with the global crisis, financial architecture or European economic fundamentals. At the core of his analysis lies the conviction that serious interventions in the labor market are the only way to trigger a satisfactory response to the competitiveness problem besetting the weaker economies in the Eurozone. He maintains that in Spain and Greece, competitive gains and improvements in current account balances have occurred through labor shedding, import contraction and export growth. Yet increasing competitiveness by reducing jobs leads to misleading results. A credible analysis should include competitiveness based “in terms of real GDP per potential worker (that is, employment plus unemployed)”.\footnote{34}

Thimann summarily switches from the “cost of capital” as a factor contributing to the reduced productivity of the peripheral eurozone countries, to the cost of the social security system and other “disincentives for investment” at the heart of the structural adjustment programmes drawn up by European creditors. He contends that “competitiveness comparisons usually focus on developments in wages and labor productivity, as cross-country differences in the cost of capital are relatively muted across developed countries”.\footnote{35} Yet this is a specious claim, given that such differences are at present far greater than those justifiable in a common currency area. In early 2014, Phil Pew, former senior ECB monetary policy official and now chief European economist at Goldman Sachs warned that “if we have to live with the current level of [credit market] fragmentation, the viability of monetary union will eventually be called into question”.\footnote{36}

Such a degree of fragmentation dramatically restricts private credit flows to the Eurozone periphery. A prime example of this is to be seen in investment in renewable energy, one of the fastest growing sectors in the global economy, and a highly capital-intensive one: according to the findings of the EU-funded “DiaCore project”, carried out by a consortium of researchers, in 2014 “the average weighted cost of capital for an onshore wind project “ranged from 3.5% in Germany to 12% in Greece”\footnote{37} Obviously, “generic country risk” and “sudden policy change risk” partly account for the fact that “it is vastly more expensive to undertake” such projects in certain EU countries than in others. However, if we take into

\footnote{34} Thimann, \textit{ibid.}, p. 146-147.  
\footnote{35} Thimann, \textit{ibid.}, p. 144.  

155
account the size and importance of the banking sector in financing the real economy in Greece, the high cost of capital creates a competitive disadvantage at least as significant as that attributed to high wages in relation to productivity, or to the relatively high cost of employer social security contributions. Even before the crisis, various sources highlighted the fact that Greek banking profitability was supported by high spreads between the rate earned on loans and that charged on deposits. A 2006 study estimated that in the previous year “net interest margins stood at 3.5%, significantly above the EU-12 average of 1.3%.” In addition, banks in Greece levy more billing fees than in the rest of the Europe, and even e-banking services have suffered from high rate charges. Invoking market forces as a way of reducing wages or social-security contributions while turning a blind eye to the high interest margin of Greek banks reveals a contradictory approach towards competition: it is taken to be healthy when rendering markets more flexible and reducing the share of labour in national income, but deleterious when it triggers a redistribution of income away from wages to profits. A non-competitive market is thus only injurious where profits suffer.

According to the Bank of Greece, the main reason for high lending rates in Greece “is that securities issued or guaranteed by the Greek state are not accepted as collateral by the Eurosystem in its regular open market operations”, even if the ECB provides access for Greek banks to the emergency liquidity assistance (ELA). This reluctance to help a struggling Eurozone economy with unprecedented rates of unemployment in recovering from a devastating chronic recession is the result of well calculated economic blackmail: only with a full implementation of labour market reforms, reductions in salaries, pensions and minimum wage, etc. will the ECB, EC and IMF show any willingness to ease the state of financial suffocation artificially imposed on Greece. One piece of evidence indicative of the way the so-called institutions work is the transcript of a teleconference between Poul Thomsen, head of the IMF’s European Department, and Delia Velculessiu, IMF mission chief for Greece, released on April 2, 2016, according to which IMF officials planned to provoke a “credit event” in the country in order to force the government to “accept our [IMF] views”, which at that juncture included pension cuts and budget savings worth 2.5% of Greek GDP.

4. The moral hazard problem and the “delinquent countries”

If EU texts at the close of the 20th century prioritized the fight against social exclusion and promoted a non-punitive approach towards poverty or academic failure and school drop-out,
an entirely different spirit prevailed in post-crisis critiques of individuals who lost their jobs, or of countries that lost access to financial markets. Creditors warned that assistance towards such individuals or countries would only come under strict conditionality clauses, which included unsustainable primary budget surpluses for many years to come. This recipe is based on a tendency in economic literature to confirm both a negative relationship between public debt and economic growth, especially when the debt to GDP ratio exceeds a certain threshold, and a positive one between high government debt and corruption. The level of government debt had become a fetish, and the lesson to be drawn from this alleged causal relationship between public debt and growth is that money should be a scarce resource controlled not by governments, but by “independent”, disinterested and purportedly socially neutral institutions such as central banks. Money should only be lent under binding terms, in order to protect society from speculation, peculation and inflation.

As they are unable to re-finance their debts by raising funds on the bond market, the smaller Eurozone countries have been subject to enhanced economic surveillance and, mutatis mutandis, have been treated like subprime mortgage holders in the USA. Subprime lending was available to people in the lowest income brackets, such as African Americans, Hispanics, women of color, single mothers etc. These borrowers were forced to pay interest averaging between $85,000 and $186,000 more than standard mortgage borrowers. One feature typical of the racial and gender discrimination bias inherent in such lending was the fact that in 2005, African American women were 256% more likely to receive a subprime mortgage than white men. Furthermore, according to a 2005 analysis of federal home lending data, African Americans, American Indians/Alaska Natives, Pacific Islanders and Hispanics relied more on

---


subprime loans than whites; borrowers of colour in almost all income brackets were three
times more likely to be offered subprime lending than their white counterparts.\textsuperscript{46} This unequal treatment based on racial-ethnic income disparities facilitated the attribution of blame to characteristics supposedly “innate” to blacks and Hispanics, such as laziness, unreliability, violent tendencies, heavy welfare dependence or unintelligence when accounting for the high share of non-performing loans in the subprime market.\textsuperscript{47} Just as the borrowers in question were held responsible for unfortunate choices based on a poor assessment of risks and future income, so the most vulnerable citizens in the Eurozone’s pariah states were accused of lacking a work ethic, enjoying welfare benefits unimaginable to other Europeans, etc. Therefore, depriving them of fundamental social and economic rights and goods was considered an appropriate punishment for their misconduct, volatility and lack of predictability.

When the US financial crisis threatened the stability of the European banking system, very few politicians and mass media outlets warned the public that the rescue packages granted to the major banks, to the tune of 1.7 trillion euro, would be an unmanageable burden on the credibility of states with high debts and deficits. But when the discussion centered on financing the public sector, debt ratios and deficits stood at the core of an international campaign against the “pariah states” of Southern Europe. With the complicity of credit rating agencies,\textsuperscript{48} which were instrumental in triggering the financial crisis, southern Eurozone countries were threatened with a permanent ban from international capital markets if they failed to submit to supervision unworthy of a sovereign state, or if they rejected the institutionalization of austerity through the Fiscal Compact and various recessionary economic adjustment programmes.

One mindset indicative of the above trend is that displayed by Charles Wyplosz, a well known French economist and consultant to the European Commission and international organizations. In March 2011, Wyplosz noted that a future system of Eurobonds should actively consider the “moral hazard aspect” of the Eurozone debt crisis, which markets are more than aware of. In this respect, the higher interest rates paid by countries “perceived as fiscally unsound act as a powerful incentive to restoring budgetary discipline”. According to Wyplosz, the mutualisation of all national government debt under the Eurobond proposal would mean each Eurozone Member State agreeing to share responsibility for debt servicing throughout the entire bloc, while at the same time determining its own separate sovereign fiscal policy. However, such an arrangement would not sufficiently address the issue of moral hazard, which can be reduced to the following question: “Why should any government exercise restraint when it knows that others will pay for its own excesses?” As a solution to this problem, he suggested that fiscally unsound countries should face disciplinary proceedings reminiscent of criminal law. The existence of [excessive] deficit in the Eurozone ought to be viewed as a crime against the Euro. Countries which persistently commit such “crimes” [“delinquent countries” as Wyplosz terms them] should be monitored and placed under surveillance just like juvenile offenders. To prevent abuses, the issuing of Eurobonds

should ex-post address the moral hazard problem created when irresponsible states gain access to cheap lending. "Thus" he concluded "solidarity comes along with some reduction in fiscal policy sovereignty".

5. Austerity as a new form of technocracy

The targets set in the context of the Lisbon Strategy underscored the close interdependence of Member States’ economies, as well as the fact that "the action (or inaction) of one Member State could have significant consequences for the EU as a whole". The need for action at the EU level in response to the debt crisis would also be an indication of the inherent links in the European economy after decades of deepening its internal market. In actual fact, however, joint action focused on inhibiting the threat Greek debt posed to European banks, rather than on easing the draconian lending conditions imposed on the troubled eurozone economies. I shall refrain from examining the widely visible effects of these choices, such as the rise of the far right in Europe, and focus instead on a less self-evident parameter: the prevalence of technocratic solutions, and the retreat of political processes and elective bodies in the formulation and implementation of economic policy in Europe.

On April 26-27, 2016, when Germany and the IMF insisted that €3.6 billion ex ante austerity measures for Greece be implemented in case of divergence from fiscal targets set by the memorandum, that is if the primary surplus for 2018 equal to 3.5% of GDP were not achieved, disagreements between the Greek government and the "institutions" concerning the constitutional legality and political expediency of ex ante implementation of contingency cuts brought the negotiations to a halt. In order to break the deadlock, the Greek government demanded a full meeting of European leaders, i.e. a “political negotiation” at the highest level. This strategy was rejected by the institutions on the grounds that it would be unacceptable to circumvent the Eurozone finance ministers, and above all, the "technical teams" from the IMF, the European Commission and the ECB.

A second instance of technocratic supremacy unfolded in December 2016, when the Greek government announced a Christmas bonus for low-income pensioners and the suspension of VAT increases on the islands then sheltering thousands of refugees. The Greek authorities maintained that the funds to cover these initiatives came out of the surplus Greece had managed to achieve, that it did not threaten fiscal targets, and that it was “up to the Greek government to distribute expenditure in the way it sees most fit and socially correct”. In contrast, the “institutions” considered the one-off payout of 600 million euro to pensioners as a unilateral decision violating the agreement between Greece and its creditors, and as a unilateral decision violating the agreement between Greece and its creditors, and as a

leveraging tool the ESM suspended short-term debt relief deal for Greece. The Greek Prime Minister sought a political solution on a three-day trip to Brussels and Berlin, where he was told point blank that such decisions lay in the hands of institutions and the Eurogroup.54 In order to break the deadlock, the Eurogroup demanded an official letter giving assurances over the provisional nature of the measure.

These incidents are indicative of the priorities set by creditors in negotiations over the implementation of austerity measures in Greece. Decision making, responsibility and enforcement are not to be entrusted to politicians, nor should economic policy be the subject of consultation, compromise or scrutiny by political parties, citizens or anyone else affected. Despite the general tendency to appoint non-elected policy experts as finance ministers - because technocrats are more likely to implement unpopular policy reforms,55 it has even been alleged that the Eurogroup is not the proper forum for such in-depth discussion.56 This line of argument is used to account for why, in the early 21st century, a broad pension reform failed in Greece. In 2013 Jesús Fernández-Villaverde, Luis Garicano and Tano Santos criticized the pre-2010 pension scheme in Greece for providing the highest “replacement rates” of any OECD country, meaning the level of pensions as a percentage of previous individual earnings at the moment of retirement.57 However, the authors studiously ignored the fact that a high replacement rate is the only way to ensure that Greek pensioners receive a decent pension, taking into account that even before crisis, salaries in the country were among the lowest in the Eurozone. Moreover, the fact that Greece was the only EU country where the minimum wage decreased after 2010 made the reduction of replacement rates even more painful for prospective pensioners.58

Arguing on the same wavelength, Thimann warns that ordinary politicians “are not familiar with intricate structural, regulatory, administrative, or social insurance issues in other countries”, and as a result the “increasing shift of policy discussions to the European political level is actually counterproductive”. Consultation will require the expertise of specialists, who ignore political cost and act with the sole aim of enhancing the efficiency and effectiveness of reforms. For these reasons, the level of negotiations must be technocratic and not national, political, or European. Emphasis should not be given to the bigger picture contemplated by elected politicians, given their “tendency to move from microeconomic issues to macroeconomic issues”. According to these perceptions, consultation over the stability support programmes at European political level shifts the center of discussion from structural reforms “that are essential to getting the economies of the crisis countries on their feet again” to macroeconomic issues such as unemployment or the sustainability of public debt. Elected

56 Thimann, Ibid.
politicians have the detrimental tendency to be “sensitive to political issues, including their reputation and chances of reelection” and thus turn direct their attention “from allocation to distribution”. 59

Distributive or expansionary policies have been held to create non "virtue growth". Fernández-Villaverde, Garicano and Santos place too much emphasis on the fact that during a credit bubble agents end up with a distorted image of market fundamentals, and in so far as “rising asset prices hide their mistakes” they are able in the short-run to “extract more rents without the fear of punishment.” Conventional analysts are not troubled by the fact that the construction of the Eurozone and financial deregulation contributed to the development of speculative bubbles. The risk of the economy overheating from an unsustainable level of consumption is invoked to invalidate policies of increasing demand and investment, or to justify wage and pension cuts; at the same time, those focusing on structural reforms rarely touch on issues of redistribution, banking regulation, capital flows and markets. Analysts who consider the negative effects of the financial bubble after the introduction of the Euro generally decouple “the bubble” from the four “freedoms”, i.e. the free movement of goods, services, labour and capital established by the Maastricht Treaty. They also dissociate abuses of the financial system from the repeal of the Glass Steagall Act, which prohibited banks from selling or underwriting insurance products. And they further disconnect the repeal of the Glass Steagall Act (1999) with USA commitments to conform to Financial Services Agreement deregulation rules, signed in 1997 under the auspices of the World Trade Organization.

Furthermore, those who emphasize the inherent imbalances generated by the establishment of the single currency remain on the fringes of prevailing narratives on the crisis. For example, Matias Vernengo and Esteban Pérez-Caldentey have pointed out that the relative increase in unit labor cost and real exchange appreciations under a regime of stable exchange rates resulted in intraregional trade imbalances in the Eurozone, which could not then be compensated for through devaluation. The lack of a “recycling” mechanism to clear imbalances from surplus to deficit countries forced the peripheral-deficit countries to implement a growth policy via increasing aggregate demand. The freedom of financial flows and enhanced liquidity “provided a false sense of prosperity” and led to private and public debt accumulation. Fiscal imbalances accumulated after 2007 as a result of increasing government expenditure for the support of banks and businesses, and a decline in revenues due to recession and decreasing production and consumption. Austerity measures, which depressed domestic demand, gave the external sector a key role to play in recovering from the crisis. But fiscal contraction would only make sense if the crisis were fiscal in nature, and not a result of internal imbalances inherent in the Euro economic model.60

Conclusion

The shift in focus from unregulated financialisation and free market abuses to the sovereign debt crisis paved the way for an even less inclusive and socially equitable model of European integration. The failure of austerity programmes imposed so as to service unsustainable public debts watered down the credibility of the connection between high levels of public debt

and austerity measures. The advocators of restrictive policies in Europe reconsider microeconomic issues, market efficiency, and labor market reforms as the proper ground for the construction of plausible arguments in favor of austerity. The introduction of new type authoritarian structures, disguised as technocratic bodies with at best indirect popular legitimacy, is likely to alienate Europe even further from its people. Such an uneven developmental strategy may well have incalculable consequences for both European societies and the rest of the world.

Author contact: kokkinakis@uoc.gr


You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-83/
The Vienna school of ecological economics


Katharine N. Farrell [Humboldt-Universität zu Berlin, Germany; Universidad del Magdalena, Colombia; Universitat Autònoma de Barcelona, Spain]

Copyright: Katharine N. Farrell 2018
You may post comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-83/

**Introduction**

Founded roughly 30 years ago, the transdisciplinary field of ecological economics is today slowly coming of age, making the publication of this forward-looking edited collection a timely contribution. Its content speaks to contemporary scientific and political discourses concerning the future of the planet earth and the place of human societies within that future. It provides an economics oriented reader with an overview of the basic concepts and issues addressed in this diverse field. Comprised of 50 short chapters, each providing a concise summary of a specific topic, the *Routledge Handbook of Ecological Economics: Nature and Society* is a good, if in places puzzling, general introduction to the social science and humanities strands of the field. Its broad scope and consistent focus on the future reflect a maturing of the field. The collection, which presents ecological economics as an alternative to neoclassical economics, is a complement to the range of more general and more conventional textbooks and handbooks already available and can serve as a helpful reference tool for teaching, theory and applied research. It is, above all, an excellent guide for those looking to move beyond conventional neoclassical economics approaches to environmental questions. Although perhaps a minor concern for the present readership, the inclusion of only one strictly ecology based contribution (out of 50 chapters), serves as a brusque reminder that there remains much work to be done before ecological economists are able to construct the epistemologically complex representations and visions of ecological economic dynamics envisioned by the discourse’s founders.

The editor has brought together an impressive list of contributors, including early ecological economists like Richard Norgaard and Joan Martínez-Alier and the journal’s current editor, Richard Howarth, and there are a number of outstanding, brief and coherent summaries of key ecological economics topics to be found in these pages. Heterodox economics enjoys a privileged position, with roughly half the chapters approaching the topic from one or a combination of heterodox economics perspective(s). These include among others, reliance on Marx, Commons and Veblen, Polanyi, Kapp and Keynes. The handbook also contains a broad range of complementary contributions approaching the theme through mixtures of political ecology, philosophy, social theory and ecological economics methodology. The scope is admirable, ranging from institutional economics (Ch.3, Vatn) and ecofeminism (Ch.5, Salleh) to the place of thermodynamics theory in ecological economics epistemology (Ch.9, Mayumi), the moral duty of present to future generations (Ch.25, Howarth) and the role of lifestyle in the perpetuation of ecologically destructive social and economic practices (Ch.15, Brand and Wissen). Readers interested in learning more about ecological economics methodology will find a helpful overview of the workings and challenges of a number of key methodologies employed in the field, including social multi-criteria analysis (Ch. 30, Greco and Munda), Q-Methodology (Ch.32, Davies) and participatory modelling (Ch.35,
Videira et al.). Those interested in learning which theories, epistemologies and ontologies guide the thinking of some of today’s leading ecological economists will find much here to occupy their attention. However, teachers and students looking for a comprehensive overview of the composition and cadres of the contemporary field of ecological economics should bear in mind that this collection presents not all but rather one among various narratives advanced within the transdiscipline.

The contributions are situated within a coherent overall structure, with the book divided into ten narratively progressive sections, each consisting of a number of ten-page chapters covering the following ten thematic areas: Foundations (2 chapters); Heterodox thought on the environment (6 chapters); Biophysical reality and its implications (5 chapters); Society, power and politics (4 chapters); Markets, production and consumption (4 chapters); Value and ethics (4 chapters); Science and society: uncertainty and precaution (3 chapters); Methods (9 chapters); Policy challenges (6 chapters); and Future post-growth society (7 chapters).

The position advanced in this collection, with conviction and often with outstanding academic rigor, is that dogmatic adherence to the ideal of economic growth, among scholars, practitioners, politicians and publics, is a driving force behind the countless socio-ecological tragedies that have characterised the close of the 20th and start of the 21st Centuries. This is a central tenant of the ecological economics discourse, which problematizes the relationship between material and energetic throughput in the human economy and material and energetic throughput in the ecological contexts within which human economies are inevitably embedded. Each chapter offers not only critiques that underpin this position but also constructive recommendations for how future research and collective action might help humanity to find a way forward that respects it.

Chapters are often cross-referenced by the contributing authors and there is a good deal of conceptual overlap between them, giving the collection a definite tone and steady narrative flow. Chapters also adhere to a standard format, making comparisons and combination of insights straightforward: first there is a general introduction to the chapter topic and to the author’s position concerning its pertinence to the field of ecological economics; next is a review of key related debates, subjects, objects and approaches; this is then followed by a section on future directions, in which authors express their opinions about how work on the chapter’s topic should be developed going forward within ecological economics; each chapter then finishes with concluding remarks and a brief list of key further readings. The standardized structure of the collection, combined with the brevity of each chapter and an excellent index, make it easy to consult and gives a definite handbook feel. However, a handbook, i.e. a text that gives clear and explicit instructions on how to apply codified procedures associated with a body of uncontested facts, it is not. Rather, this is a work of advocacy, worthy of admiration on those grounds, and certainly well worth consulting but nonetheless, a work of advocacy.

**Advocacy, realism and issues of normativity**

In the Preface, in Chapter 1 and through the inclusion of a number of contributors not typically associated with the core ecological economics discourse, Spash has constructed a case in favour of a particular approach to the field of ecological economics. The general thrust of this case, reflected in the collection’s two closing sections, ‘Policy challenges’ and ‘Future post-growth society,’ is to lay out a suite of analytical and methodological foundations for
interpreting ecological economics as a post-growth / degrowth alternative to conventional, mainstream, neoclassical economics. That case is based mainly in contemporary heterodox economics theory and is framed at the start of the collection, and in several chapters, through reference to Bhaskar’s Critical Realism. It adopts a standpoint that is not universally accepted among ecological economists, nor, in my view, entirely compatible with the original aims and contemporary challenges of the transdiscipline. This is a subtle point, and so I wish to be clear. Taken as a contribution toward the necessarily complex, non-reductive, sometimes dishevelled community of propositions and approaches that comprise the transdiscipline of ecological economics, the position advanced in this collection addresses a core project: developing concrete alternatives to a continued reliance on proxy monetary-value based representations of the economic contribution of ecological phenomena (Farrell, 2009[2005]; 2007; Farrell and Silva-Macher, 2017). However, it would be inappropriate, in my view, to consider it the definitive word on the topic.

Similarly, with respect to the ecological aspect of ecological economics, among fifty contributions, only five directly address biophysical dynamics, while, among these, content is related mainly to economics and epistemology. In the only contribution approaching the topic squarely from an ecology perspective (C.12, Devictor), the author presents a well-established critique of reductionist, or equilibrium, ecology theory and highlights ways in which reductionist ecology and reductionist economics reinforce one and other. He fails, however, to take into account the importance of far-from-equilibrium thermodynamics in the work of many ecological economics scholars, not least among them the ground breaking socio-ecologist C.S. Holling and the founder of the bioeconomics discourse, Nicholas Georgescu-Roegen. A closer look at Devictor’s chapter, “The Biophysical Realities of Ecosystems,” both in terms of its content and in terms of Spash’s decision to include it in this collection, raises two issues that I believe to be of great importance to the field of ecological economics:

1) to what extent does contemporary academic work concerning environmental issues continue to be biased by a euro-descendent colonialist cosmology of conquest and control that dismisses the collaborative humans-within-nature perspective of first nations in favour of the confrontational humans-against-nature perspective that underlies the European projects of industrialisation and Enlightenment?

2) on what basis is it possible to ensure epistemological rigor in the production of epistemologically complex, inter- and transdisciplinary work concerned with the state and future of the planet earth and the place of humans in them?

With regard to the first point, Devictor presumes, in his chapter, as does Spash in his, and as do many other, although not all, of the contributors to this collection, the existence of something they refer to as ‘Nature’, which is presumed to exist beyond the reach of human understanding and is fixed and immutable in its reality. This is a Platonic presumption, repeated in the work of Aristotle and later confronted, in the context of social science, by Weber. It implies not only that there are right and wrong ways to make sense of the ecological economic Gestalt (Farrell and Silva Macher, 2017) with which I am in full agreement, but also that the monistic euro-descendent epistemology that presumes an a priori distinction between real and not-real, is universally applicable.

Leaving aside the question of whether or not other epistemologies are convincing, their exclusion from consideration, in place of dogmatic adherence to the primacy of this Platonic presumption leaves ecological economics hollow. In place of responsible recognition of the normative implications of actively engaging the hermeneutically complex challenge of
supporting contemporary humanity in reshaping a globalised life-world, we are left with normativity masquerading as objectivity. The impossibility of achieving such objectivity is, in contrast, a central epistemological proposition of the complexity theory that informed the work of almost all proto-ecological economists (Prigogine, 1997; Cilliers, 2005).

With regard to the second point, concerning epistemological rigor – and I understand the two to be related – since methodologies operationalise and inform theory, closing down methodological options also implies closing down theory options. In Devictor’s chapter, the far-from-equilibrium thermodynamics assumptions of dynamic indeterminacy in the works of Georgescu-Roegen, Boulding and Holling are left aside, along with the epistemological indeterminacy that they imply. Interestingly, it is precisely these epistemological implications of far-from-equilibrium thermodynamics assumptions, disregarded by Devictor, that occupy Mayumi’s (Ch.9) attentions in what is a genuinely ground breaking intervention and an absolute must read. This heterogeneity of contributions can be understood as a symptom of the interdisciplinary context of ecological economics, where there is a constant risk that the quality of crossover contributions will be accepted uncritically, as would appear to have been the case with Devictor’s chapter, when editors and team members are confronted with interventions that reference discourses with which they are not familiar. Methodological pluralism and epistemological openness can help to mitigate this by creating an intellectual climate that invites and facilitates participation of experts from a wide range of disciplines, increasing the likelihood that a team has recourse to reviewers sufficiently familiar with the discipline in question, be it anthropology or biochemistry, to mount competent critiques of discipline specific interventions.

**Handbooks, context and the purpose of Ecological Economics: the importance of standpoints**

In fairness to Spash, the idea of a ‘handbook of ecological economics’ is perhaps itself an oxymoron, as the epistemological foundations of the field, rooted in the rise of modern complexity theory during and after the Second World War, militate against the type of codifying one would expect from a handbook. This peculiarity of the field, and of this collection, provides an opportunity to reflect both on the place of normativity in ecological economics – a topic directly addressed in several of its contributions, and on the ethical, epistemological and methodological challenges associated with making that normativity transparent – a topic that figures in only a handful of them. In recent years, ecological economics, first established at the close of the 1980s as an upstart transdiscipline running against the grain, has become increasingly accepted as a key intellectual space for working through the questions of what is sustainability and how might it be realised (Faber, 2008). This has brought newcomers to the discourse whose positions are not tightly bound to its origins and there has arisen, within the journal *Ecological Economics*, a debate regarding the discourse’s purpose (Nadeau, 2015). The editor of this collection has been actively engaged in this debate (Spash, 2012; 2013): is it a sub-field of economics? A proto-discipline requiring refinement? A broad-church transdiscipline encompassing ecology, economics and a range of other disciplines, which should remain methodologically agnostic? A social movement? something else altogether? While these are questions that can only be answered by contemporary ecological economists, it behoves a thoughtful scholar to reflect upon the views of those who have opened the path along which one is travelling.
Notwithstanding their inevitable flaws, laid bare over time, looking back to the critical works of early, proto-ecological economists from the 1960s and ‘70s, such as Kenneth Boulding, Nicholas Georgescu-Roegen, Howard T. Odum and Donella Meadows, two consistent themes can be discerned:

1) there is an epistemological imperative to include attention to both economic and ecological phenomena in the construction of non-reductive, complex representations of ecological economic dynamics

2) choosing to do so reflects not only a scientific but also a normative, ethical position.

The question as to how point 1) might be addressed was taken up as a central matter of concern in the first issue of the journal *Ecological Economics*, in 1989, where, in keeping with point 2), two countervailing, explicitly normative arguments, endorsing alternative strategies were openly discussed. Costanza (1989) argued in favour of developing transdisciplinary synthetizing methodologies that combine ecological and economic analyses in new and productive ways, while Norgaard (1989) argued in favour of what he called 'methodological pluralism,' warning against shutting down debate and closing off avenues of enquiry in an effort to achieve the appearance of formalism in the study of a profoundly complex subject.  

Spash, the editor of this collection, is an outspoken advocate of formalising and standardizing the field’s methodology using Critical Realism as a structuring frame. While it may appear pedantic, I would argue that maintaining both synthesis and pluralism squarely on ecological economics’ methodological agenda is not a question of style but of epistemology: is there one correct way to conceptualise and envision the dynamic relationship between ecological and economic processes? This is what Farrell and Silva-Macher (2017) call the ecological economic *Gestalt*. Given that normative reflection upon the constitution of the human/environment relation is a basic feature of the human condition, does the topic’s inherent, irrevocable epistemological indeterminacy oblige the ecological economist to tolerate a degree of intellectual diversity and an underdetermined ontology that resists resolved formalisation?

Arguably, advocacy of a unitary formalisation of the ecological economic approach, even one as ostensibly loose as Bhaskar’s Critical Realism, reflects a standpoint built up from within the Euro-descendent culture of the Scientific Method, Logical Positivism and Analytical Philosophy, where there is one real world and the task of science is to discover it. Advocacy of a less closed position, where the Euro-descendent approach to making sense of humans’ place in the world can be complemented by post-modern and non-modern epistemologies is about ‘staying with the trouble,’ as Haraway (2016) puts it. It is a standpoint favoured by myself and many other ecological economists, including some of the contributors to this collection, as far as I can discern. It is a standpoint that, in keeping with the general tone of this collection, sees an alternative to the European model of development (Acosta, 2013). That is, one that rejects the hegemonic Euro-descendent monetarist economy model that equates Gross Domestic Product with human wellbeing; one that rejects the idea that there is a single path of human progress – from hunter gather, to herder, to farmer, to industrialist, to post-industrialist – embracing, instead, the idea of a “pluriverse, [which] signals struggles for bringing about, worlds and knowledges otherwise – this is, worlds and knowledges constructed on the basis of different ontological commitments, epistemic configurations, and practices of being, knowing and doing” (Escobar, 2012[1995]: xxviii).

---

1 For an economics oriented review of these origins, see Nadeau, 2015.
Without wishing to suggest that this collection’s Euro-descendent bias is intentional, its presence, like the dearth of ecology contributions, highlights the insidiousness with which European ideations of progress and nature continue to limit the thinking of scholars and activists in both the minority (cash rich) and majority (cash poor) worlds. By defining ecological economics as a means for upending the neoclassical economics paradigm, and for refuting the associated politics that takes economic growth as an end desirable in itself, this collection paints itself into a corner. Its identity is determined through reference to that which it is not, and in this way still intimately bound to that definitional context. For example, the language of many of the collection’s contributions reflects a general developmentalist (Escobar, 2012[1995]) bias: e.g., in Chapter 15, where Zaman refers to ‘pre-market societies’ as a stage in the process of human social development, implying that contemporary human societies where barter, reciprocity and gifting continue to define socio-economic relations are somehow developmentally retarded.

In addition to this linguistic and narrative bias, the composition of authors also reflects an extreme bias, based in Europe or the Euro-descendent world (80 % of all authors, 95% if the USA is included), with almost half of all the first authors (48%) coming from either Austria, the United Kingdom or the United States (listed in order of predominance). Were this collection entitled, in a geographical sense, The New Austrian School of Ecological Economics, the presence of such strong bias would be less disconcerting. However, these Euro-descendent and Austrian institutional biases are not immediately obvious, as the author list includes neither the institutional affiliations nor the personal biographies of the contributors. I looked many of them up in order to prepare this review.

Furthermore, in spite of a rich epistemological and methodological diversity reflected in its overall composition, the collection begins with a section entitled Foundations, consisting of two highly normative, epistemologically closed interventions. The first, from the editor, argues that there is a pressing need to formalise the methodological foundations of the field of ecological economics, which he explicitly casts as an alternative to neo-classical economics, situated within a Marxian analytical tradition. The second is a sparsely referenced harangue against the methodological paucity of contemporary ecological economics written by two authors without reputation in the field, arguing that it lacks impact because it is unstructured and that Bhaskar’s Critical Realism should be adopted as the overarching approach, in order to correct that failure. Their utter disregard for the innumerable ecological economics texts addressing questions of methodology over the past thirty years, and for the contemporary renaissance in that dialogue currently underway in the journal Ecological Economics is, frankly, embarrassing. Notwithstanding their location, in the section of the book labelled ‘Foundations,’ these chapters provide the reader, not so much with information regarding the foundations of the ecological economics discourse, as with argumentation advocating that Critical Realism should be adopted as the epistemological foundation for contemporary ecological economics.2

Conclusion

Looking to the future, and on this I am in full agreement with Spash and commend him and his contributors for engaging with that challenge, the bias reflected in his selection of

---

contributors provides a wonderful backdrop for opening up discussion regarding the epistemological and methodological consequences of a latent Euro-descendent bias that continues to persist within ecological economics. There are many academically outstanding interventions brought together in this collection, which will provide the critical experienced ecological economist with much food for thought, and the ecological economics novice with a helpful introduction to an important strand of the wider discourse, one we might refer to, perhaps, in keeping with tradition, as the Vienna School of Ecological Economics.

References


Author contact: katharine.farrell@qub.ac.uk

SUGGESTED CITATION:

You may post and read comments on this paper at [https://rwer.wordpress.com/comments-on-rwer-issue-no-83/](https://rwer.wordpress.com/comments-on-rwer-issue-no-83/)
Board of Editors

Nicola Acocella, Italy, University of Rome
Robert Costanza, Australia, The Australian National University.
Wolfgang Drechsler, Estonia, Tallinn University of Technology
Kevin Gallagher, USA, Boston University
Jo Marie Griesgraber, USA, New Rules for Global Finance Coalition
Bernard Guerrien, France, Université Paris 1 Panthéon-Sorbonne
Michael Hudson, USA, University of Missouri at Kansas City
Anne Mayhew, USA, University of Tennessee
Gustavo Marqués, Argentina, Universidad de Buenos Aires
Julie A. Nelson, USA, University of Massachusetts, Boston
Paul Ormerod, UK, Volterra Consulting
Richard Parker, USA, Harvard University
Ann Pettifor, UK, Policy Research in Macroeconomics
Alicia Puyana, Mexico, Latin American School of Social Sciences
Jacques Sapir, France, École des hautes études en sciences sociales
Peter Söderbaum, Sweden, School of Sustainable Development of Society and Technology
Peter Rafford, USA, The Radford Free Press
David Ruccio, USA, Notre Dame University
Immanuel Wallerstein, USA, Yale University

EDITORS: Edward Fullbrook,
ASSOCIATE EDITORS: Jamie Morgan
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=_NDUzNDkzMHlwzMzIzODZ8cGFlX25ld3NAYmRpbnRelcm5ldC5jb218MTA5OTU4_