

real-world economics review

- Subscribers: **24,614** Subscribe [here](#) [Blog](#) ISSN 1755-9472
- A journal of the World Economics Association (WEA) 13,183 members, [join here](#)
- Sister open-access journals: [Economic Thought](#) and [World Economic Review](#)
- back issues at www.paecon.net recent issues: [67](#) [66](#) [65](#) [64](#) [63](#) [62](#) [61](#) [60](#) [59](#) [58](#)

Issue no. 68, 21 August 2014

In this issue:

Monetary policy in the US and the EU after quantitative easing Thomas Palley	<u>2</u>
Processes vs. mechanisms and two kinds of rationality Gustavo Marqués	<u>10</u>
A systems and thermodynamics perspective on technology in the circular economy Crelis F. Rammelt and Phillip Crisp	<u>25</u>
Back where we started from: 'the Classics' to Keynes, and back again Roy H Grieve	<u>41</u>
Demand theory is founded on errors Jonathan Barzilai	<u>62</u>
The central bank with an expanded role in a purely electronic monetary system Trond Andresen	<u>66</u>
Financialization, income distribution and social justice: Recent German and American experience Robert R Locke	<u>74</u>
Recovering Adam Smith's ethical economics Thomas R. Wells	<u>90</u>
The human element in the New Economics Neva Goodwin	<u>98</u>
Placing economists' analyses of antidumping in an antitrust context John B. Benedetto	<u>119</u>
Board of Editors, past contributors, submissions and etc.	<u>147</u>

from the Blog: [American public sector jobs under 6 presidents](#) - [World student movement could become major player](#) - [Bundesbank call s for higher wages across Eurozone](#) - [The revolt of \(part of\) the top 1% of the top 1%](#) - [US median wealth is down by 20 percent since 1984](#) - [Unsound on equality](#) - [How the Goldman Sachs guys wrecked Europe](#) - [Polanyi's methodology in the Great Transformation](#) - [Re-thinking the Definition of "Public Goods"](#) - [What to do to make economics a relevant and realist science](#) - [Can economists explain much?](#) - [Low interest rates?\(3 graphs\)](#) - [Keynes on the use of mathematics in economics](#)



You may join the WEA for free [here](#)

Special RWER issue on Piketty's *Capital* coming late September

Contributors

Dean Baker, Víctor A. Beker, David Colander, Ping Chen, Edward Fullbrook, James Galbraith, Jayati Ghosh, Claude Hillinger, Steve Keen, Merijn Knibbe, Richard Koo, Richard Parker, Heikki Patomäki, Ann Pettifor, Alicia Puyana, Jeffery Sachs, Jacques Sapir, Lars Pålsson Syll, Yanis Varoufakis, Robert Wade

Monetary policy in the US and EU after quantitative easing: the case for asset based reserve requirements (ABRR)¹

Thomas I. Palley [Washington, D.C, USA]

Copyright: Thomas I. Palley, 2014
You may post comments on this paper at
<http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

Abstract

This paper critiques the Federal Reserve's quantitative easing (QE) exit strategy which aims to deactivate excess liquidity via higher interest rates on reserves. That is equivalent to giving banks a tax cut at the public's expense. It also risks domestic and international financial market turmoil. The paper proposes an alternative exit strategy based on ABRR which avoids the adverse fiscal and financial market impacts of higher interest rates. ABRR also increase the number of monetary policy instruments which can permanently improve policy. This is especially beneficial for euro zone countries. Furthermore, ABRR yield fiscal benefits via increased seignorage and can shrink a financial sector that is too large.

Keywords Quantitative easing, asset based reserve requirements, exit strategy

JEL reference E52, E58

1. Introduction

This paper critiques the Federal Reserve's quantitative easing (QE) exit strategy which aims to deactivate excess liquidity via higher interest rates on reserves. That strategy is equivalent to giving banks a tax cut at the public's expense, and it also risks domestic and international financial market turmoil. The paper offers an alternative exit strategy based on asset based reserve requirements (ABRR) which avoids the adverse fiscal and financial market impacts of higher interest rates. Implementing a system of ABRR also increases the number of monetary policy instruments which can permanently improve policy. This is especially beneficial for euro zone countries. Furthermore, ABRR yield fiscal benefits via increased seignorage, and they can also help shrink the financial sector which many believe has become too large owing to financialization of the economy.

2. What is QE?

QE is an unconventional monetary policy used by central banks when standard monetary policy has become ineffective because the central bank's short-term policy nominal interest rate is at or near zero and cannot be lowered further to stimulate economic activity. It involves the central bank buying financial assets (like mortgage backed securities and collateralized debt obligations) from commercial banks and other financial institutions, and thereby increasing the monetary base.

¹ This paper was presented at the Progressive Economy Forum held at the European Parliament, Brussels, Belgium, 5-6 March, 2014.

3. The effects of QE

QE in the US has had four major effects. First, it has significantly expanded the size of central bank balance sheets. Thus, the Federal Reserve's balance sheet has expanded by over \$3 trillion, rising from \$920 billion at the end of December 2007 to over \$4.2 trillion at the end of February 2014.

Second, a concomitant part of that expansion has been an increase in excess reserves of the commercial banking system which as of February 2014 were \$2.5 trillion.

Third, QE has lowered long-term interest rates and increased stock prices and financial wealth, which are the channels whereby it has stimulated real economic activity.

Fourth, QE has contributed to significant capital inflows to emerging market (EM) economies. That is because EM economies have had far higher interest rates over the last several years than the US and other developed economies which have been pursuing QE policies.

4. The challenge of exiting QE

The US economy has now healed significantly since the financial crisis of 2008 and the Great Recession of 2009. The Federal Reserve has therefore begun to implement its strategy for exiting from QE and normalizing monetary policy in anticipation of more normal future economic conditions.

Exiting QE raises three major challenges. First, how should monetary policy deactivate the excess reserves of the banking system so that they do not finance either future unwanted inflationary private sector expansion or destabilizing asset price bubbles?² Second, how should policy avoid triggering asset price disruptions (i.e. a stock market or bond market crash)? Third, how should policy avoid causing exchange rate disruptions from international capital flow reversals that could trigger financial market turmoil in the rest of the global economy?

5. The Federal Reserve's exit strategy

The Federal Reserve's current strategy involves three components. The first component is so-called "tapering" when the Federal Reserve scales back its QE purchases, gradually reducing them to zero. Thereafter, the Federal Reserve's balance sheet will be gradually reduced by having the private sector redeem maturing securities held by the Fed, which will reduce private sector holdings of reserves.

The second component is transparent forward guidance which is intended to enable markets to plan for the QE exit process and thereby help stabilize financial markets.

² The Federal Reserve needs to deactivate excess reserves because it needs to re-establish reserve scarcity in order to push the future short term policy interest rate above zero. Additionally, it needs to deactivate reserves if there are any monetarist transmission channels whereby excess liquidity on bank balance sheets changes bank portfolio and lending behavior.

The third component is a plan to pay an increased interest rate on reserve deposits at the Fed. This is intended to give banks an incentive to maintain their excess reserve deposits with the Fed, thereby deactivating them.

6. The Fed's strategy is already not working

Unfortunately, there is every indication that the plan is already going wrong. First, talk of tapering in August 2013 contributed to immediate (albeit short-lived) stock and bond market price declines.

Second, the anticipation and beginning of tapering may have contributed to international capital flow reversals and exchange rate disruptions that have already rocked emerging market (EM) economies.

Third, the payment of higher interest rates on excess reserves promises to be very expensive. It is also expansionary, which runs counter to the purpose of raising interest rates. The expense is very clear. Given banks hold \$2.6 trillion in total reserves, every one hundred basis point increase in interest rates costs the Federal Reserve \$26 billion. If the Fed's policy interest rate returns to 3 percent, that would cost \$78 billion. That is an effective tax cut for banks because the Fed would pay banks interest, which would reduce the profits it pays to the Treasury. The banks, which were so responsible for the financial crisis, would therefore emerge winners yet again. Taxpayers, who bailed out the banks, would once again bear the cost.

Paying interest to banks would also run counter to macroeconomic policy purpose since it would be pumping liquidity into the banks when policy is explicitly trying to deactivate liquidity. That smacks of policy contradiction.

7. Asset based reserve requirements (ABRR): an alternative exit strategy

Implementing a system of ABRR offers a superior alternative QE exit strategy (Palley, 2010). Not only can such a system of address the exit challenge more effectively and more cheaply to taxpayers, it also offers long-term policy benefits regarding the conduct of monetary policy. In particular, it addresses the problem of asset bubbles that caused the financial crisis. Furthermore, ABRR have particular benefits for Europe because they can help address the loss of national monetary policy resulting from the creation of the euro.

7.a) What are ABRR?

ABRR consist of extending margin requirements to a wide array of assets held by financial institutions (Palley, 2000, 2003, 2004). ABRR require financial firms to hold reserves against different classes of assets, with the regulatory authority setting adjustable reserve requirements on the basis of its concerns with each asset class.

ABRR are easy to implement, use the tried and tested approach of reserve requirements, and would fill a major hole in the existing range of financial policy instruments. However, maximum effectiveness of an ABRR approach requires system-wide application to diminish

possibilities for avoidance that can contribute to instability. For instance, if applied only to banks, ABRR would encourage lending to shift outside the banking sector, thereby promoting shadow banking that has been shown to be relatively less stable. To fully succeed, reserve requirements must therefore be set by asset type, not by who holds the asset.

ABRR are also compatible with existing regulation but they differ from conventional notions of quantitative regulation. The historic focus of bank regulation has been the prevention of bank runs and the traditional form of regulation has been reserves on liabilities (deposits). ABRR have a different focus which is avoiding excessively risky speculation, and they can also help shrink a bloated financial sector resulting from financialization.³ This focus on curbing risky speculation is similar to capital standards. However, as discussed further below, ABRR are more flexible than capital standards because they impose reserve requirements rather than equity requirements and they are also counter-cyclical rather than pro-cyclical.

7.b) ABRR provide a superior exit strategy from QE

ABRR can provide a superior exit strategy from QE. Instead of paying increased interest rates on excess reserves, the Federal Reserve would impose a reserve requirement on assets with the aim of mopping up the excess liquidity QE has created. Banks would have to hold reserves against their assets, and so too would other financial institutions.

How does this help? Instead of raising interest rates to deactivate liquidity, the Federal Reserve would impose asset reserve requirements. That avoids paying interest to banks and rewarding them. It also avoids implicitly penalizing taxpayers by lowering the Federal Reserve's profits, and thereby reducing the profits it pays to the Treasury.

Imposing reserve requirements on assets will mean loan interest rates rise to compensate lenders for money tied up backing those loans. However, it will have less of an effect on loan interest rates than increasing the central bank's policy interest rate. If the loan interest rate (i_L) is a mark-up (m) over the central bank's policy interest rate (i_F), the loan rate without ABRR is given by:

$$(1) i_L = i_F + m$$

The loan rate with ABRR is given by:

$$(2) i_L = [1 + k]i_F + m \quad 0 \leq k < 1$$

k = asset reserve requirement. The loan rate increases because lenders charge borrowers for the liquidity they must hold against loans. A higher asset reserve requirement (k) increases the loan rate (i_L), as does a higher policy rate (i_F). As regards exiting QE, the goal is to absorb excess reserves held by banks. Imposing a reserve requirement that impacts all loans absorbs more excess reserves with less impact on the loan rate than does raising the central bank's policy rate. The latter only affects reserves by discouraging lending at the margin.⁴

³ For details about financialization and its economic effects see Epstein (2001), Hein (2012), Krippner (2005), and Palley (2008, 2013a).

⁴ Raising the policy interest rate induces banks to hold more as reserves by reducing the amount they loan (L). The approximate effect on banks' reserve demand (R) is $dR/di_F = |dL/di_L \cdot di_L/di_F|$. Raising the reserve requirement on loans causes banks to hold more reserves by compelling them to back all loans with reserves (kL). The approximate effect on reserve demand is $dR/dk = L + kdL/di_L \cdot di_L/dk$.

Consequently, an ABRR strategy will impose less disruption on the overall economy. The specific effects on bond and stock markets would depend on the particulars of how reserve requirements were assessed. The stock market would likely strengthen if stocks were assessed with a zero reserve requirement while bonds had a positive requirement. This is because stocks would become relatively more attractive compared to bonds. Conversely, stock prices would likely drop if stocks were subjected to a positive reserve requirement and bonds were zero-rated.

Lastly, deactivating excess liquidity via ABRR diminishes the likelihood of capital flow reversals from EM economies back to the US. That is because financial capital would have a reduced incentive to flow back to the US given the short term policy rate (i_F) is unchanged. Indeed, imposing ABRR might even cause some US outflows by financial capital seeking to avoid reserve requirements.

7.c) ABRR would improve the conduct of monetary policy

Most importantly, ABRR can permanently improve the conduct of monetary policy (Palley, 2000, 2003, 2004). There is widespread recognition that the financial crisis which triggered the Great Recession was significantly due to financial excess, particularly related to real estate. Moreover, there is growing recognition that the real estate bubble was just another, albeit the largest, in a string of bubbles.

The toleration of serial bubbles over the past two decades reflects profound intellectual failure among central bankers and economists who believed inflation targeting was a complete and sufficient policy framework (Palley, 2005). It also reflects lack of policy instruments for directly targeting financial market excess. With central banks relying on the single instrument of the short-term interest rate, using that interest rate to target asset prices would be like using a blunderbuss that inflicts massive collateral damage on the rest of the economy.

ABRR offer a simple solution to this problem by providing a new set of policy instruments that can target financial market excess, leaving interest rate policy free to manage the overall macroeconomic situation. By obliging financial firms to hold reserves, the system requires they retain some of their funds as non-interest-bearing deposits with the central bank. The implicit cost of forgone interest must be charged against investing in a particular asset category, reducing its return. Financial firms will therefore reduce holdings of assets with higher reserve requirements and shift funds into other lower-cost and thus relatively more profitable asset categories.

By adjusting reserve requirements on specific asset classes, central banks can target specific financial sector imbalances without recourse to the blunderbuss of interest rate increases. For example, if a monetary authority was concerned about a house price bubble generating excessive risk exposure, it could impose reserve requirements on new mortgages. This would force mortgage lenders to hold some cash to support their new loans, raising the cost of such loans and cooling the market.

If a monetary authority wanted to prevent a stock market bubble, it could impose reserve requirements on equity holdings. This would force financial firms to hold some cash to back their equity holdings, lowering the return on equities and discouraging such investments.

ABRR also act as automatic stabilizers. When asset values rise or when the financial sector creates new assets, ABRR generate an automatic monetary restraint by requiring the financial sector come up with additional reserves. Conversely, when asset values fall or financial assets are extinguished, ABRR generate an automatic monetary easing by releasing reserves previously held against assets.

In all of this, ABRR remain fully consistent with the existing system of monetary control as exercised through central bank provision of liquidity at a given interest rate. They are also compatible with the existing regulatory system based on capital requirements, liquidity requirements, and liability based reserve requirements (i.e. reserve requirements on deposits). However, they are superior to these systems because ABRR are a form of financial automatic stabilizer. That is the opposite of capital requirements which are a form of financial automatic de-stabilizer. Equity capital tends to be destroyed in economic downturns when it is hardest to replace, and requiring firms to come up with more capital to cover losses and deteriorated asset quality deepens downturns. The reverse holds in booms when capital standards can contribute a pro-cyclical dynamic.

At the microeconomic level, ABRR can be used to allocate funds to public purposes such as inner city revitalization or environmental protection (Thurow, 1972; Pollin, 1993). By setting low (or even negative) reserve requirements on such investments, monetary authorities can channel funds into priority areas, much as government subsidized credit and guarantee programs and government-sponsored secondary markets have expanded education and home ownership opportunities and promoted regional development.⁵ Conversely, ABRR can be used to discourage asset allocations that are deemed socially counterproductive.

ABRR also promise significant fiscal benefits by increasing seigniorage revenue for governments at a time of fiscal squeeze. To the extent that required reserves constitute a tax on financial institutions, that tax is economically efficient given the costs of resolving financial crises. It will also shrink a financial system that many believe is bloated.

7.d) Advantages of ABRR for the euro zone

ABRR are especially attractive for the euro zone. That is because they can help address the instrument gap created by the euro's introduction. The euro's establishment has required member countries to give up their own interest rate policy and exchange rates. That has reduced the number of policy instruments, creating problems for country economic policy management. ABRR can fill this policy instrument gap because they can be implemented on a geographic basis by national central banks.

Property lending, which has been a major focus of concern, is particularly suited to this. If the euro zone were suffering excessive house price inflation, the European Central bank (ECB) could raise reserve requirements on mortgage loans secured by property. In addition, national central banks could have the power to set reserve requirements above (but not below) the rate established by the ECB. Thus, if Spain or Ireland were suffering excessive house price inflation, their national central banks could raise reserve requirements on mortgage loans secured by property in those countries. That would raise mortgage loan rates in Spain and Ireland without raising rates in other countries.

⁵ Negative reserve requirements would work by giving banks a reserve requirement credit that could be applied against other asset holdings with positive reserve requirements.

Nationally contingent ABRR within the euro zone would create some incentive to shop for credit across countries. That means ABRR will work best when linked to geographically specific assets that cannot evade the regulatory net. This includes secured loans, particularly mortgage loans which are secured by collateralized property. However, even with jurisdictional shopping, ABRR will still be effective. That is because jurisdictional shopping is costly and that shopping cost creates space for some degree of cross-country interest rate differentials that a system of ABRR can take advantage of.

Additionally, ABRR can be used to encourage holdings of euro zone country government debt by assessing a low or negative reserve requirement on such assets. Such a measure would have helped greatly in the euro zone financial crisis of 2010–2012. A negative reserve requirement would have increased demand for government debt, thereby discouraging speculative attack against euro zone government debt and lowering bond interest rates. That might have prevented government bond interest rates from spiking and causing Europe's public debt crisis.

7.e) ABRR as a means of tackling financialization

Finally, ABRR offer an important policy instrument for addressing the problem of financialization and reducing the size of the financial sector (Palley, 2013b). Epstein (2001, p.1) defines financialization as referring to “the increasing importance of financial markets, financial motives, financial institutions, and financial elites in the operation of the economy and its governing institutions, both at the national and international level.”

Palley (2013a, p.9) describes financialization as transforming “the functioning of economic system at both the macro and micro levels. The principal impacts are to (1) elevate the significance of the financial sector relative to the real sector; (2) transfer income from the real sector to the financial sector; and (3) increase income inequality and contribute to wage stagnation.” These adverse effects of financialization are now being recognized by mainstream economists and Cechetti and Kharroubi (2012), from the Bank of International Settlements, report that too large a financial sector lowers growth.

The adverse macroeconomic effects of financialization call for shrinking the size of the financial sector. ABRR can play an important role as part of a strategy to do so by imposing reserve requirements on those parts of the financial sector that have expanded excessively. That can lower returns in those activities, thereby shrinking them.

8. Conclusion: the question of policy authority

This paper has argued that ABRR provide a superior exit strategy from QE compared to the Federal Reserve's current proposed strategy of paying interest on reserves. Not only would an ABRR based exit strategy be cheaper and more effective, it would also yield significant improvements in the conduct of monetary policy by giving the Federal Reserve new policy instruments to target specific financial sector disruptions. An ABRR strategy would also yield similar benefits to the ECB and the euro zone. The Federal Reserve already has the legal authority to impose ABRR on commercial banks. If it needs new legal authority to enable it to impose ABRR beyond the banking sector, it should seek that authority from legislators. Lack

of authority is not an argument against ABRR: instead, it is an argument for new legislation granting authority.

References

Cechetti, S.G. and Kharroubi, E. (2012) "Reassessing the impact of finance on growth," BIS Working Paper No. 381, Bank of International Settlements, Basel, Switzerland.

Epstein, G. (2001), "Financialization, rentier interests, and central bank policy," manuscript, Department of Economics, University of Massachusetts, Amherst, MA, December.

Hein, E. (2012), *The Macroeconomics of Finance-dominated Capitalism – and its Crisis*, Cheltenham, Edward Elgar.

Krippner, G. (2005), "The financialization of the American Economy," *Socio-Economic Review*, 3, 173 - 2008.

Palley, T. I. (2000), "Stabilizing finance: The case for asset based reserve requirements," *Financial Markets and Society*, The Financial Markets Center, Philomont, VA, August.

Palley, T. I. (2003), "Asset price bubbles and the case for asset based reserve requirements," *Challenge*, 46 (May – June), 53 – 72.

Palley, T. I. (2004), "Asset based reserve requirements: Reasserting domestic monetary control in an era of financial innovation and instability," *Review of Political Economy*, 16 (January), 43 – 58.

Palley, T. I. (2005), "A Post Keynesian framework for monetary policy: Why interest rate operating procedures are not enough," in Gnos, C., and L.-P. Rochon (eds), *Post Keynesian Principles of Policy*, Cheltenham, E. Elgar.

Palley, T. I. (2008) "Financialization: what it is and why it matters," in Hein, Niechoj, Spahn, and Truger (eds.), *Finance-led Capitalism? Macroeconomic Effects of Changes in the Financial Sector*, Metropolis – Verlag, Marburg, 29 - 60.

Palley, T. I. (2010) "The politics of paying interest on bank reserves: A criticism of Bernanke's exit strategy," *Challenge*, 53 (May/June), 49 – 65.

Palley, T. I. (2013a), *Financialization: The Macroeconomics of Finance Capital Domination*, Macmillan/Palgrave, Basigstoke, UK.

Palley, T. I. (2013b) "Making finance serve the real economy," in Palley, T.I. and Horn G. (eds.), *Restoring Shared Prosperity: A Policy Agenda From Leading Keynesian Economists*, CreateSpace.com, 2013.

Pollin, R. (1993), "Public credit allocation through the Federal Reserve: Why it is needed; How it should be done," in Dymski, G.A., Epstein, G., and Pollin, R., (eds.), *Transforming the U.S. Financial System: Equity and Efficiency for the 21st Century*, Armonk, NY:M.E.Sharpe.

Thurow, L. (1972), "Proposals for re-channeling funds to meet social priorities," in *Policies for a More Competitive Financial System*, conference proceedings for Federal Reserve Bank of Boston, June, 179 – 189.

Author contact: mail@thomaspalley.com

SUGGESTED CITATION:

Thomas I. Palley, "Monetary policy in the US and EU after quantitative easing: the case for asset based reserve requirements (ABRR)", *real-world economics review*, issue no. 68, 21 August 2014, pp. 2-9, <http://www.paecon.net/PAERreview/issue68/Palley68.pdf>

You may post and read comments on this paper at <http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

Processes vs. mechanisms and two kinds of rationality

Gustavo Marqués [University of Buenos Aires, Argentina]

Copyright: Gustavo Marqués, 2014
You may post comments on this paper at
<http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

Introduction

The contribution of standard mainstream economics to the development of the global financial crisis that began in 2007 has been analyzed in several recent papers. Some authors have denied that mainstream economic theory has had any responsibility in its occurrence and have argued that it could not have been predicted or anticipated. Colander (2010) argued that the actual economic system is so complex that available current models (“DSGE models”) are not sophisticated enough to detect the imminence of such a crisis. In fact, according to him, the situation is even worse because it would not have been possible to foresee the crisis even if the profession had been provided with more sophisticated mathematical models. The problem lies then not in standard economic theory *per se* but in the nature of economic phenomena.

In opposition to this interpretation, some authors have pointed out that the prevailing view of financial systems has much of the responsibility for the occurrence of the crisis. Haring (2013), to quote a recent testimony, argues that conventional textbooks in the field offer a distorted treatment of the nature of money and the process of its creation, and hide the close relationship between central banks and private banks. In particular, the current literature does not mention the fact that most of the money created comes from private commercial banks¹. Even more troubling is the hypothesis of efficient markets, which in conjunction with rational expectations theory makes impossible (by hypothesis) the occurrence of crises, inducing a wrong analysis of the role of liquidity and ignoring the conditions that can lead to a crisis. Standard economics were not well fitted to appraise the health of the global economy.

“Why did the majority of economists fail to foresee the Global Financial Crash of 2008?” We can now answer this question quite simply as follows. The majority of economists were members of the Neoclassical School. They therefore accepted the core of the neoclassical paradigm, namely equilibrium theory.

According to equilibrium theory, a market governed by free competition moves into a Pareto-optimal equilibrium. Now the financial markets throughout the world had in the decades before 2008 been deregulated and so made to approximate to a freely competitive market. Hence neoclassical economists deduced from their economic paradigm that these financial markets would move towards equilibrium rather than crashing in a catastrophic fashion. (Gillies, 2012, p. 32)

¹ Such omissions posse a problem, because “this rhetoric frames the minds of central bankers, other policy makers, academics and - through economic journalists educated with the same textbooks - the general public, in a very unfortunate way. This prevents them from understanding the current financial crisis and from drawing the right policy conclusions from it” (Håring, 2013, p. 3).

The usual “naturalistic” perspective of current mainstream philosophy of economics, with its concern for “recovering” the rationality of mainstream modeling practice (in order to illuminate the problem each model was attempting to solve) seems to be at best irrelevant in such circumstances, and, at worst, a part of the network of concealment. What is required is a different way of approaching the economy and its theoretical representations. We need to rethink the economy. The aim of the present paper is to contribute to this task. It joins the wave of growing interest – shown in numerous recent articles – in building an approach able to pay more attention to real economies and the actual economic practice. A non-mainstream epistemological and philosophical vision about how to look at the economic problems is now under construction. The proposals in this regard are still partial and piecemeal, but exhibit important common grounds around issues of common interest. This program contrasts with the exaggerated emphasis that the mainstream philosophy of economics has placed on the analysis of arbitrary economic models, and contests the still ungrounded claim that unrealistic “parallel” worlds have nonetheless cognitive relevance.

Rethinking economics is an eminently philosophical task. It is not the same as (or not reduced to) making economic theory, something that has to be done by the economists themselves. It is rather to think about both economic theory and the real economies. Rethinking real economies includes the examination of their particularity as a process and their ontological characteristics; rethinking economic theory includes assessing the epistemic and ontological assumptions of current theoretical practice (Fullbrook, 2013, 2014). Both tasks are urgent and necessary, given that the mainstream philosophy of economics pays no attention to real economies and is inclined to justify (a-critically) any sort of conventional theoretical practice. Rethinking economics must be sensitive to the growing awareness that economics as a discipline should focus on solving important problems, which immediately affect the lives of most people (poverty, unemployment, growth, inflation, to mention just those which require more urgent attention), rather than finding sophisticated solutions to theoretical puzzles which only restrictive academic circles find worth of interest. I encourage people to ask Boland’s question any time a suspicious model is exposed (Boland, 2005)².

In a recent work (Ivarola et al, 2013), some central ideas the authors believe should be part of a more adequate philosophy of economics have been outlined. It was held that one of the fundamental aspects of the new vision is to reject the idea that the discovery of mechanisms is or should be the central goal of economic theory. It is better to consider economics as providing inputs in order to think (and command) processes. In this paper, four main aspects that have not been addressed – or are not properly highlighted – in that work are discussed. They are the following.

² “I am, as you can imagine, a methodological terror in my department, and have been for twenty years. I am a terror not because I come in and pontificate about methodology. I never talk to my colleagues about Popper or anything – they couldn’t care less, and I understand that. I understand my audience. All I do is, when they give a seminar on whatever fancy thing they are doing at the time, I will ask them, before they get started, why did they bother to do this paper? Now this is a terrifying question for people. First of all, they spent \$50,000 or \$100,000 on research, and they know they’ve got garbage, and they don’t want you to let anybody know *that* because they are responsible for the research, and so on. You know, the worst thing you can have is somebody asking the question of why you are doing this” (Boland, 2005, p. 157)

1) Wishful thinking. Emphasizing processes, particularly their openness, could be interpreted as if some form of wishful thinking is promoted. This is not the view I sustain. I rather defend a limited social constructivism, where rigidities (constraints) make a difference and not all one imagines or wants is an attainable target.

2) Openness and stability. That processes are “open”, in our sense, implies nothing about their stability. Given that the recent financial crisis has put on the agenda the discussion of complex systems, it is important to examine some of the conditions that led to instability in the global economy and the difference in nature between this feature and the open character of economic processes. The ability to integrate both views in the same analysis is also examined.

3) The existence of mechanisms. Recognizing economic processes does not eliminate the possibility that in some circumstances they may be working together with more rigid sequences of events, which may be assimilated to mechanisms (in one of the senses considered by the current literature about this issue). The nature of these more stable sequences as well as the particular conditions that allow their manifestation must be investigated and understood.

4) Rationality. It is difficult to accommodate the traditional view of rationality within a framework in which the presence of uncertainty and the openness of economic processes is admitted. However, this can be done if we leave off the deliberative and ex-ante notion of rationality, one which precedes decision making, and develop a concept of ex-post rationality, which consists in validating the decisions already adopted.

Properly understood, these points are compatible with (and complementary to) the processual view of economics that I wish to defend. In what follows I will examine the contribution of some recent works to approach these issues and the way in which these ideas can (or cannot) be articulated with my perspective.

1. Management of processes does not imply wishful thinking (it is not true that anything works)

White (2013) discusses the anti-crisis policies adopted by the United States and central European countries during 2008-2011. He shows that what he calls “ultra easy monetary policy” implemented by the central banks of these countries during that period was doomed to fail. To sustain this position White appeals to the idea that our current actions can have inevitable consequences in the long term, which are unexpected and possibly unwanted. Such policies, according to White, are not a “free lunch”: they provide some immediate benefits but impose very high costs in the future. His analysis has interesting implications for the distinction between economic processes and mechanisms, a distinction that I want to defend. White bases his analysis on a classical formulation of the thesis:

“No very deep knowledge of economics is usually needed for grasping the immediate effects of a measure; but the task of economics is to foretell the remoter effects, and so to allow us to avoid such acts as attempts to remedy a present ill by sowing the seeds of a much greater ill for the future” (Ludwig von Mises, *The Theory of Money and Credit*).

Other authors have also supported this perspective (Hayek and Popper among them). Although this argument has been offered within the framework of a liberal and philosophically conventional view of the economy, many economists (probably most), even within the heterodox field, agree with it. This thesis, if it has any sense, has to be interpreted literally. It says that in some cases when a measure X of economic policy is adopted with the purpose of solving urgent problems, it may lead, in the long term, to consequences Y that can worsen the very issues that demand immediate attention. In such cases doing X inexorably leads to Y. And this result will occur regardless of subsequent intentions and actions of the authorities that have produced X. In addition, it is claimed that the unwelcomed consequences are only unexpected to the untrained observer. As the quote from Mises suggests, a sort of cognitive asymmetry between experts and laymen is assumed. Furthermore, it is argued that the task of economics as a science, is (or should be) the ex-ante identification of the causal link between the initial policy measure and its remote consequences, a nexus that lies beyond the reach of common citizens. To the extent that the causal link is known, the consequences of the action may be anticipated. This has been part of the battery of arguments used against any interventionist project on the market. White's contribution seems to be on the same track. His paper investigates whether, despite its proven failure it would have been possible to implement policies of easy credit avoiding the already known consequences. His findings seem to support an anti-interventionist vision.

“Stimulative monetary policies are commonly referred to as ‘Keynesian’. However, it is important to note that Keynes himself was not convinced of the effectiveness of easy money in restoring real growth in the face of a Deep Slump. This is one of the principal insights of the General Theory. In the current circumstances, two questions must be addressed. First, will ultra easy monetary conditions be effectively transmitted to the real economy? Second, assuming the answer to the first question is yes, will private sector spending respond in such a way as to stimulate the real economy and reduce unemployment? It is suggested in this paper that the answer to both questions is no” (White, 2013, p. 23).

However, White does not oppose all monetary policy *per se*, and he is not trying to defend a general argument against interventionism. Rather, he examines the particularly extreme form that this policy adopted during the recent crisis. The text suggests that once an extreme decision as the one which is the subject of analysis has been taken (interest rates close to 0 and very easy access to credit allocation), certain undesirable effects will ensue. It sounds reasonable: extreme measures may limit to such extent the margin of maneuver of the authorities, that it will be very difficult or impossible to counteract or rectify their effects later. But the article does not rule out that adopting more cautious policies, potential adverse effects could have been tempered or eliminated by subsequent measures. It is then possible to intervene more intelligently, and it is this idea that the author seems to convey in his final conclusion.

“Looking forward to when this crisis is over, the principal lesson for central banks would seem to be that they should lean more aggressively against credit driven upswings, and be more prepared to tolerate the subsequent downswings. This could help avoid future crises of the current sort. Of course the current crisis is not yet over, and the principal lesson to be drawn from this paper concerns governments rather more than central banks. What central banks have done is to buy time to allow governments to follow the

policies that are more likely to lead to a resumption of 'strong, sustainable and balanced' global growth. If governments do not use this time wisely, then the ongoing economic and financial crisis can only worsen as the unintended consequences of current monetary policies increasingly materialize" (White, 2013, p. 50).

It is worth noting also that White's results against monetary policies are obtained by examining the operations of the market in the traditional manner, considering only the action of purely economic factors. Limiting the analysis to the domain of economic variables, and given the extreme nature of the measures, his paper shows that no sustained growth could be achieved in these conditions. But it should be noted that the alleged inevitability of the crisis is achieved by leaving out consideration of the possible introduction of regulations and substantial institutional changes able to force agents and other economic actors to adopt different behaviors. The introduction of regulations and institutions opens new scenarios. One of the subsequent measures that could be adopted is the enforced condition that the money provided to the private sector be used to generate productive employment activities, preventing its speculative use³. White does not examine the effect of a regulation of this kind. He does not, because he seems to be analyzing the operating market system in a "pure" form. There is little doubt that, in the absence of regulations, the market is unable to prevent speculative bubbles and subsequent crises.

In my opinion, it would be more appropriate to argue that his article shows the inadequacy of wishful thinking, not interventionism. White opposes wishful thinking with the epistemological tools he finds at hand (the theory of unintended consequences, which presupposes the existence of transmission mechanisms). This commitment is effective for his purpose, but is too strong, and unnecessary. To oppose that view, it is enough to show that there are concrete rigidities (physical and institutional restrictions) that cannot be dismissed. As pointed out, the extreme measures he considers generate constraints that can disable any further managerial effort to redirect the economy in order to achieve the objectives initially pursued. Our perspective rejects wishful thinking, and at the same time is not committed to the (ubiquitous) existence of transmission mechanisms, known ex-ante by economic science and ignored by laymen and "men of practical knowledge"⁴. A very needed task of the "new paradigm" for rethinking the economy should develop a stronger criticism of the alleged inevitability of certain outcomes and a more sustained effort in showing the conditions under which such "inevitable" are obtained.

Another interesting aspect of White's analysis is that he assigned to the thesis of the unintended consequences of our current decisions a meaning that differs from the traditional sense: instead of pointing out the supremacy of the "experts" (in this case, users of mainstream models) compared with non-trained people, he argues that those models have NOT been effective in showing the unintended consequences of the extreme liquidity policy implemented before, during and after the crisis. White turns against mainstream economic theory a key part of mainstream epistemology.

"The unexpected beginning of the financial and economic crisis, and its unexpected resistance to policy measures taken to date, leads to a simple conclusion. The variety of economic models used by modern academics and

³ For a discussion of the distinction between production and speculation, see Reinert, 2013.

⁴ See Ivarola et al, (2013).

by policymakers give few insights as to how the economy really works. If we accept this ignorance as an undesirable reality, then it would also seem hard to deny the possibility that the policy actions taken in recent years might also have unintended consequences” (White, 2013, p. 30).

Here’s an interesting argumentative twist. White believes that standard models do not help us understand how the economy works and those who were surprised by the upcoming of unexpected consequences this time were the users and builders of these models. Other authors have stressed the same point. Not surprisingly, if policies are implemented on a wrong basis they will probably show unintended consequences. We are at the antipodes of Mises’ claim: while “best” available models failed to anticipate the long run, many “outsiders” have accurately perceived far-reaching consequences (Keen, 2013). Our interpretation is that “outsiders” have been more effective in this task because they have paid more attention to real economies and conceived them more as social processes than as economic mechanisms, and this conceptual change allowed them to focus on factors that were omitted in the conventional picture.

One point of disagreement between my vision and White’s point of view concerns the nature of the unexpected economic consequences. In his opinion, they come from epistemological (or gnoseological) reasons. I consider them differently, although both views seem to be complementary with one another. Though we do not know what will happen in the long run, this ignorance is not a pure epistemological phenomenon but has ontological foundations. It is the open nature of the process which often makes the resulting end unexpected. And it is not caused by bad models. There is no hope that better models may prevent us from being surprised by the future outcomes of our present policies. Radical uncertainty prevails.

Although given uncertainty there is no scientific way to assign relative probabilities to different possible outcomes, agents and actors can know for sure what outcomes among the feasible ones would be most beneficial for them in case they occur. They are experts in their own interests. Therefore all relevant actors seek to manage the results! In fact, some big players make extensive use of a rhetorical device, insisting that they know what cannot be known: saying again and again that such and such policies will/will not yield some desired/undesired results; but their own haste to assert such a thing again and again shows just the opposite (they have to fight in several ways, including discursive procedures, in order to be sure that their favorite results ultimately be realized).

2. Instability and openness

In Helbing and Kirman (2013) the notions of complexity and instability are linked. I will show the points of agreement and disagreement between my view and their position. The first thing to be noted is that my analysis centers on the “openness” of economic processes, while Helbing and Kirman focus on their “instability.” Although the two notions can be connected, they are indeed different. The “openness” originates in the insurmountable cognitive limitations of agents, ultimately grounded on ontological reasons, while “instability” that characterizes “complex dynamic systems” comes basically from two key features: a) excessive interconnection between the elements of the system (in this case, banks or financial firms); b) presence of big elements, with so substantial economic volume that they generate what the authors call the “too big to fail problem”. A system that has these properties

(like the current financial system in the dominant countries) can lead to undesirable domino and cascade effects.

The necessity of a new framework to approach economics has been demanded by Fullbrook (2013, 2014) and by Helbing and Kirman, when they refer, respectively, to the formation of “New Paradigm Economics” (NPE) and “New Economic Thinking”⁵. Our notion of an intervenible, open-ended process based on expectations does not incorporate the two above mentioned features (see Ivarola et al, 2013). But it might do so. And to show how both concepts (openness and instability) can be articulated would be an important contribution to the construction of a new philosophical approach to economics.

In our analysis, two sources for openness of economic processes are considered:

- a) Agents’ expectations
- b) Actors’ interventions

Although we acknowledge that changes in expectations can drive the process to different final states (even stagnation), we have not emphasized the phenomenon of instability. Actually, our exposition does not reveal the particular form the recent financial crisis has taken (nor the phenomenon of bubble formation). In particular, we have not shown (or mentioned) that the several ways in which changes in expectations can influence the process crucially depend on the type of relationships between the components (agents) of the system; and we have not remarked that the size of some of these agents may be central to the outcome of the whole process. This is an important aspect of the economic processes that has been correctly captured by Helbing and Kirman.

Although the notions of instability and openness are compatible and within real economies both phenomena are actually present, they are different and should be distinguished for analytical and practical purposes. Even if, as the authors suggest, to design a stable system were possible (one where no opportunities for undesirable effects, like domino or cascade, can occur), our analysis would still be worthwhile, in the sense that the process would remain open-ended and would be based on intervenible expectations.

Another important difference is that they do not distinguish, as we do, between agents and actors⁶. However, taking actors into account may be important for approaching a broad set of issues. In Ivarola et al (2013) the role of actors in the formation of expectations was pointed out, but we have not emphasized enough their influence in the construction of the relevant context (regulations, laws, institutions) that provide the frame in which those expectations are formed.

Actually, an important source of instability comes from actors’ interventions, even from those that are not agents in the strict sense of the term: communication media, corporations, unions and the like. Considering actors in such a broad sense may illuminate other sources of instability beyond those coming from firms and relations among them. Particularly, including

⁵ Several authors share this goal. Lawson has been advocating for a long time for a specific alternative (critical realism), that takes into account the ontological nature of social phenomena. Also Hodgson (2012) encourages the broadening of the conventional economic view incorporating psychological and institutional aspects, while Dow (2012a, 2012b) proposes an alternative pluralistic approach to mainstream vision of economic theory.

⁶ This distinction is traced between *roles* played by economic subjects. Agents and actors are not mutually exclusive classes. A subject may belong to both of them as long as he performs both roles.

actors into the analysis allows the introduction of politics and conflicting interests into the approach to economic processes. It might be helpful for examining the kind of political and economic instability that we find in underdeveloped countries.

Actors' participation also enriches the discussion about forms of uncertainty in actual economies. Expectations are not something agents somehow form by themselves, but are based ultimately on the struggle among the actors in order to shape the context and influencing agents' expectations. How this confrontation will end is uncertain for the actors, a fact that reinforces the openness of economic processes. Uncertainty reappears here in a second level.

Conflictuality and uncertainty are the clue to understand why actors are not mere observers but active participants striving for influencing the outcome. Taking as given that stability benefits most economic agents, enforcing ways to stabilize these processes is most difficult because even when adequate knowledge of institutions is available, preventing swings hurts the interests of powerful speculators.

Perhaps the point at which our analysis diverges most from that offered by Helbing and Kirman is that they seem to believe that the main tool to stabilize the economic system is to acquire better and more "scientific" knowledge of its operations. They argue that rules and stabilizing institutions must be "tested" before being implemented. We welcome such tests, in case they are possible, but we think that, even if those tests were successfully conducted, the transition from experiment to practice faces two main problems. First, that of external validity (not yet satisfactorily solved). Second, and more importantly, we doubt that having well tested theories and models is necessary to implement the needed policies. A pre-scientific knowledge of both human behavior and the performance of institutions may be sufficient to start. Throughout the process there are always opportunities for corrections and for further action. In any case, none of these decisions are based on decisive "proofs" or tight valid demonstrations. This point will become clearer later when we introduce the notion of validating rationality.

3. Existence of transmission mechanisms

According to the dominant view of financial markets, access to easy credit driven by central banks in both the U.S. and the developed countries of the European Union, should have led to economic growth. The transmission mechanism is the following: availability of easy credit (at rates of zero or near zero interest) will raise the price of stocks (including real estate within this category), generating a wealth effect that will in turn increase consumption and the GDP. See some testimonies.

"Before the current turmoil began, Federal Reserve Chairman Ben Bernanke's hope was that rising asset prices would lead to a 'wealth effect' that would encourage the American consumer to start spending again, and thus help the American economy finally leave the 'Great Recession' behind" (Keen, 2013, p. 3).

Alan Greenspan has been even more explicit.

“...the stock market is the really key player in the game of economic growth... The data shows that stock prices are not only a leading indicator of economic activity, they are a major cause of it. The statistics indicate that 6 percent of the change in GDP results from changes in market value of stocks and homes” (Quoted by Keen, 2013, p.3).

The statements of both Fed officials were not capricious or arbitrary, but were founded on the “conventional wisdom” provided by what were supposed to be the best available models about the operation of the financial system. Very schematically, according to these models, the structure of the mechanism is the following:

+Δ credit → +Δ prices of stocks → +Δ wealth → +Δ consumption → +Δ GDP

However, the implemented measures did not yield the expected results, finally generating the crisis. A more detailed picture of the facts that includes crucial information for understanding how the bubble is formed and why it is so precarious is this: easy access to credit implies borrowing from those who take it. Taking debt allows firms to make investment transactions that exceed their real estate (“leverage effect”), buying shares and raising their prices with their decision, then sell them with a return, and finally undertake a new operation of purchase and subsequent sale. Here is a synopsis of this alternative mechanism:

+Δ credit → +Δ debts → Leverage effect → +Δ buying assets → +Δ prices of assets → +Δ returns → +Δ buying assets → +Δ prices of assets → +Δ returns → Bubble

Scheme 1

Herein is the bubble: as demand for shares holds high their prices continue to rise and this opens up new business opportunities. However, the increase in stock prices is unrelated to any increase in the economy (GDP), which in the final period reached a 12:1 ratio! (Keen, 2013, p. 5.) Given the gap, the danger grows that at some point potential buyers may consider that such prices are unsustainable and can only move downwards. There begins the deflationary period (post-bubble). To make matters worse, an indebted society like that of the U.S.A, has a small portion of their personal income available for consumer spending. Consequently, it will fail to recover enough (Hudson, 2013). Against conventional wisdom that greater liquidity makes financial markets work better, many recent papers point out that the belief in this false idea, as well as the monetary policies it suggested were the triggering cause of the sequence of events that led to the crisis.

How much of “real” mechanisms and how much of open processes are these so called “economic mechanisms”? It is an important issue, because the more the processes resemble mechanisms the less relevant management activity becomes. If economic theory, qua science, presupposes the idea that underlying transmission mechanisms are working, the political management of the economy tends to become useless. As mentioned, trained economists are likely to model such mechanisms. It is important, therefore, to investigate the conditions that must be met for the occurrence of this type of regular event sequences. The texts to which we refer in this paper show some of them.

a) Extreme measures. White's paper provides an illustration of the consequences of extreme monetary policies. Loans at zero or near to zero interest rates are unable to stimulate the economy, whatever the authorities do in later stages. However, White suggests that more moderate policies could be managed successfully.

b) Absence of constraints imposed on banks and credit-takers. The granting of loans was not subject to reasonable restrictions. On the one hand, there was a poor discrimination about who qualified as creditworthy and who did not. On the other hand, loans were not restricted to a particular kind of use meant to enhance economic activity. In these circumstances, credits were used in speculative activities that did not generate employment.

c) Lack of regulations on the use of the owned resources, as well as on the size of companies and the kind of relations they can keep with each other. The colossal size of some companies and banks gives them a significant market power, transforming those markets in which they operate in imperfectly competitive markets. Moreover, some mega-firms have multiple connections with other companies whose conjoined capital and asset stocks largely exceed the total product of many countries. These circumstances give them a formidable extortive power, leading to the "too big to fail problem": authorities cannot let these firms go bankrupt because that would drag behind them many of the productive and financial systems.

Some consequences of points (a) to (c) have already been discussed. Here I want to focus on the social nature of socioeconomic sequences and the role of an ubiquitous crucial condition that allows the emergence of economic mechanisms. It has been argued that the presence of regulations and institutions *hinders* the goals pursued by individuals. For example, to quote a favorite example of Popper's (1996), consider the simple situation faced by a pedestrian trying to cross a highway. His options will be very different whether there is or not a pedestrian path. The path forces him to cross only at the designated place. This suggests that the number of choices available to the pedestrian appears to be linked positively with the absence of regulations, while their presence would impose limits on their movements across the road. Because the rules are in force an observer might anticipate where a pedestrian would cross to the other sidewalk. Including rules help to construct simple models of pedestrian behavior. Enforcing constraints to limit the set of choices an individual faces is a procedure that can be found both in real life and in conceptual representations (models).

It is not often noted, however, that a model may focus on a typical course of action (which is privileged over other possible sequences of events) incorporating a particular type of *de*-regulation. Indeed, starting with self-centered individuals, many theoretical economic models, representing transmission mechanisms (more complex and detailed versions of so-called "economic laws"), get these sequences of allegedly unavoidable events assuming the *absence* of (some type of crucial) regulations.

Let's see how it works. Full respect for private property means *absence of regulations that restrict the use of owned resources*. Obviously, if someone is self-centered, has access to a clearly optimal choice and there are no restrictions for taking it, he will certainly choose that option (whether or not it is collectively beneficial). You can then build a model that shows that, given certain changes in the situation agents will inevitably choose *this* option. A transmission mechanism emerges, which, in the manner of Mises, is presented as inevitable. No mention

is made that one of the premises of the alleged inevitability is the freedom given to the agent to do as he pleases with his property.

If given an extreme situation (like a flood or an earthquake) nothing prevents me from increasing the price of an essential commodity that I have for sale, being self-centered I will substantially raise the price. Here is a “mechanism” (increasing demand leads to price increases), asserted by the standard economic knowledge, which works in the case of natural disasters. However, if regulations against such behavior were enforced, the predicted increase would be prevented⁷.

Schematic representations, as illustrated in Scheme 1, seemingly regular mechanisms, are merely man-made sequences, which could be avoided if appropriate regulations designed to hinder collectively undesirable behaviors are taken into account. Scheme 1 is a diagram of a truncated open process (what is depicted is just one of the many branches of the tree of feasible sequences of events). There is no natural (inevitable) connection between self-centered agents and speculative activities.

No attempt is made here to address the complex question of which specific activities are collectively beneficial and which are not. That is the subject of further analysis. The core of my argument is that many sequences of events that are presented as *mechanisms* (i.e., as sequences of events organized in a stable way and leading to results known beforehand) in theoretical models are actually socially constructed by the presence (often tacit) of regulations and institutions that eliminate otherwise alternative options. My argument is against the alleged *naturalness* of social sequences modeled within theoretical models. These sequences do not reflect social laws (like physical laws), or mechanisms in the usual sense of the term (used in current mechanistic literature). When they are represented within theoretical models, they are not much more than modeled representations of truncated processes, which are open-ended in reality. Theoretical mechanisms are obtained assuming as “natural” and given (i.e., unchangeable as a matter of principle) institutional features that are actually historically determined and perfectly modifiable.

Someone may find it foolhardy to suggest that the principle of individual freedom for disposing of their own resources could be removed (and maybe should be challenged in particular circumstances). However, modern societies are characterized by the ubiquitous presence of restrictions upon individual behavior. Many sets of laws and regulations within a society have this limitative purpose. More importantly, the presence as well as the absence of restrictions on private property is strictly a social issue (i.e., the product of social conventions adopted on the basis of the existing balance of forces at the time regulations and norms are sanctioned). My main claim here is that conventional economic mechanisms are obtained assuming explicitly or tacitly crucial de-regulations (especially on big firms and mega-firms’ activities) as part and parcel of the sequences of events that generate regular economic phenomena. This happens both on models and on real economies.

The above considerations may be useful to address phenomena such as bubbles and the volatile nature of the current financial system. The absence of measures to regulate the market structure, especially the net of business connections, as well as the concentration of resources and the use that can legitimately be made of owned goods and capital, leaves the

⁷ It is interesting that the city of New York has legislation that prevents this type of self-centered behavior, regulating the prices of a set of highly needed goods with the arrival of winter.

door open to economic behavior with negative impact on growth and employment. To the extent that, as some authors argue, there can be a distinction between productive and speculative activities, what is required is the creation of institutional instruments to expand and facilitate the first type of activities and temper or prevent those of the second type (Reinert, 2013). This seems to require an extension of public law over private law.

4. Validating rationality

Rosenberg (2014) added his views to those that oppose the predominant way of modeling in economics and advocate the necessity of trying a different approach to social issues. Moreover, he focused not on the examination of representations, but in the analysis of economic processes themselves, which reaffirms the primacy of ontology over epistemology (a perspective supported by Lawson). This is the key to reversing what Fullbrook (2014) designated as “upside-down economics”. Although I agree with Rosenberg’s critique of the extreme rationalism that pervades mainstream economic theory and share his claim that there are few opportunities for the uprising of regularities in human affairs I disagree in several aspects with his treatment of both points. First, he approaches social phenomena with tools taken from the natural sciences (now it is not physics but biology that provides the relevant paradigm). Moreover, in his approach deliberate action and intentionality, as such, does not play any role. From their point of view, to understand the endurance of economic relations all that is needed is the reception they get from the “environment” (the utility economic institutions provide to their “consumers”). The characteristic traits of economic processes would then be the combination of unexpected changes (transformation or creation of economic institutions) and subsequent agents’ response to those changes. Rosenberg’s perspective is similar to that described by Alchian (1950), in which individuals do not adapt to the environment but are rather adopted by it. From this perspective there is no gain in distinguishing between agents and actors and no role for purposeful actions. Finally, while the presence of uncertainty is highlighted on Rosenberg’s account, its only purpose is to dismiss the Rat-choice approach, not to point out a crucial characteristic feature of economic processes. Besides, he seems to believe that assuming uncertainty there is no room for rationality in the analysis of social phenomena.

My proposal seeks to put in the center of the analysis the specificity of social processes, while rescuing a role for deliberate rational action in economic processes. In particular, as was argued in section 3, social regularities can be deliberately constructed. Moreover, although Rosenberg is right in pointing out that much of human behavior is based on emotions and heuristics (“rules” in words of Keynes, 1937), the fact remains that individuals do use rational strategies designed to influence economic processes. These strategies are what I call *validating rationality*. In what follows I will briefly describe what this kind of rationality is. But first, let’s see an approach to rationality I reject because of its incompatibility with the existence of uncertainty.

Deliberative and ex-ante rationality

Savage’s paradigm of decision making under uncertainty incorporates the main features of this type of rationality. It is assumed that the subject does not know what state of the world is actually in place, but he knows all the possible states and what the results of the various actions that may be undertaken would be if the world were in any of these states. Suppose, for simplicity’s sake, there are only two states, S1 and S2, and three actions a1 to a3.

S	A	C
	a1	c1
S1	a2	c2
	a3	c3
	a1	c4
S2	a2	c5
	a3	c6

The subject does not know whether the world is at S1 or S2 states, but he knows that if it is S1 and he does action a1, the result that will be obtained is c1, while if the world is at S2 he will get c4. The exercise is repeated for the other alternatives. This model of decision making is based on three basic assumptions:

- 1) Ignorance of the prevailing state of the world;
- 2) Impotence to alter that state;
- 3) Knowledge of the laws of nature (i.e., the connections between actions and outcomes given the state of the world is supposed to be known).

Point (1) makes it explicit what the individual is uncertain about. He is not uncertain about all the possible states (that is known). What he does not know is which of them is actually in place. From (1) and (3) it follows that he does not know either which outcome he will get when performing an action. This model aims to clarify how and in what sense it might be said that an action undertaken before knowing what the prevailing state of the world is can be considered rational. I call it *an ex ante model of deliberative rationality*. To behave rationally all that is required is that subjects be consistent with their (subjective) beliefs about the states of the world and their expectations about the results they can get by acting in a certain way. In other words, to be rational is to select the prospect (lottery) that gives him higher expected utility. Perhaps, in the end, the subject discovers that the world was not as he thought and that, therefore, the resulting consequences were not what he had imagined, but this does not detract at all from the rationality of his initial decision.

Validating rationality

If we give up the assumption (2) that agents have no power over the states of the world, and consider that there is a time interval between the action taken and the outcome, other options become available for them. An individual (or a firm) can know what the impact on the world will be of *other* actions, taken after what we call here the “decision” in the strict sense has been adopted. Let’s suppose a lapse of time divided into two periods, $t_0 - t_1$ and $t_1 - t_2$. In t_0 a subject takes the decision a1, assuming the prevailing state of the world along $t_0 - t_2$ will be S1 (the one in which c1 is expected). But then, he does not stay idle, but undertakes some

additional actions b_1, \dots, b_n , designed to produce (or help to create) the needed state S_1 . These actions, additional to (and successive of) the initial decision, are aimed at the transformation of reality in a precise way in order to get the desired result. We may call them *validating actions*.

A good example of validating action is propaganda, which tries to install at the top of the agents' preferences a product whose production has already been decided (or has already been finished). Let's suppose that a bad decision has been taken. For instance, to produce a doll that nobody wants (something that is seen after the production process has come to an end). One solution is to change the preferences of the agents for this product. Changing preferences is an important way to change the states of the world. Remember that this was not a possibility within Savage's paradigm.

Propaganda is just an instance of a much wider repertoire of validating strategies. If a company decides to manufacture a car able to reach high speeds, it can try to ensure the existence of norms that will add value to the differential feature of its product. For instance, once its decision has been taken the company may decide to become engaged in aggressive lobbying in order to influence the sanction of new traffic rules that will be in force when the product reaches the market. The creation of new institutions is something alike to creating a complementary good for the one that is being sold.

Perfect competition models, which assume that agents lack market power, are in harmony with the *deliberative* and *ex ante* conception of rationality. The individual or the firm, making some calculations, take a decision at an initial time t_0 and then cross their arms and wait to see what destiny has reserved for them when the time lapse ends. In a real, not perfectly competitive market, however, there is scope for considering a different notion of rationality. Mega-companies undertake actions to reduce the competitiveness of potential rivals. They can implement practical actions and strategies aimed at creating favorable conditions for obtaining the desired results. An agent is rational not merely because he is consistent or because he is capable of performing a maximizing calculation once his interests or priorities have been defined, but because he *knows how to meet his expectations*. And he *successfully meets them*. Agents' rationality is not now of the deliberative and *ex-ante* type, but rather of the *practical* and *ex-post* type. Rationality cannot consist in a mere deliberative process that ends when you have chosen the best option beforehand. Not under uncertainty. There are no good (or best) options that can be recognized *ex-ante*. *Good choices must be constructed*. The rational enterprise strives to transform reality in a way that validates its previous decisions. The rationality of the decision originally adopted as such is revealed only at the end of the process. The paradigm of business rationality is self-made prophecy (perhaps wishful thinking made real by self-effort).

References

- Alchian, A., (1950), "Uncertainty, Evolution, and Economic Theory", *Journal of Political Economy*, LVIII, February – December, 1950.
- Boland, L., (2005), *Critical Economic Methodology: A Personal Odyssey*, London and New York: Routledge, p. 157.
- Colander, D., (2010), "The economics profession, the financial crisis, and method", *Journal of Economic Methodology*, Vol. 17, No. 4, December 2010, 419–427
- Dow, S., (2012a), Different Approaches to the Financial Crisis, *Economic Thought*, Volume 1, Issue 1.

Dow, S., (2012b), *Foundations for New Economic Thinking, A Collections of Essays*, New York, Palgrave Macmillan.

Fullbrook, E., "New paradigm economics", *real-world economics review*, issue no. 65, 27 September 2013, pp. 129-131,

Fullbrook, E., *New Paradigm Economics versus Old Paradigm Economics Interview with Edward Fullbrook, Conducted by Paul Rosenberg*, *real-world economics review*, issue no. 66, 13 January 2014, pp. 131-143.

Gillies, Donald, *Economics and Research Assessment Systems*, *Economic Thought* 1: 23-47, 2012.

Häring, Norbert, "The veil of deception over money: how central bankers and textbooks distort the nature of banking and central banking", *real-world economics review*, issue no. 62, 25 March 2013, pp. 2-18.

Helbing, Dirk and Kirman, Alan, "Rethinking economics using complexity theory", *real-world economics review*, issue no. 64, 2 July 2013, pp. 23-52.

Hodgson, G., "On the Limits of Rational Choice Theory", *Economic Thought*, Volume 1, Issue 1, 2012

Hudson, Michael, "From the bubble economy to debt deflation and privatization", *real-world economics review*, issue no. 64, 2 July 2013, pp. 21-22.

Ivarola, L., Weisman, D., and Marqués, G., "Expectations-Based Processes. An interventionist account of economic practice". *Economic Thought*. Vol. 2, N° 2, 2013; pp. 30 – 42

Keen, Steve, "A bubble so big we can't even see it", *real-world economics review*, issue no. 64, 2 July 2013, pp. 3-10.

Keynes, J. M., (2002), (first Ed.,1936), *The General Theory of Employment, Interest and Money*. <http://www.marxists.org/reference/subject/economics/keynes/general-theory/ch12.htm> 1937, "The General Theory of employment", *The Quarterly Journal of Economics*, February 1937.

Marqués, G., 2014, *Notas para la construcción de una teoría práctica de la acción económica* (<http://ciece.com.ar/documentos-en-elaboracion-gustavo-marques>)

Pettifor, Ann, "The next crisis", *real-world economics review*, issue no. 64, 2 July 2013, pp. 15-20,

Popper, K., (1996), *The Myth of the Framework*, Edited by M. A. Notturmo. London and New York, Routledge.

Pullen, John, "Diagrammatic economics. A review of Blaug, M. and Lloyd, P. eds. *Famous Figures and Diagrams in Economics*, Cheltenham, U.K.: Edward Elgar, 2010, pp.xvii, 468. ISBN 978 1 84844", *real-world economics review*, issue no. 65, 27 September 2013, pp. 20-29,

Reinert, Erik S., "Civilizing capitalism: "good" and "bad" greed from the enlightenment to Thorstein Veblen (1857-1929)", *real-world economics review*, issue no. 63, 25 March 2013, pp. 57-72.

Rosenberg, Alex, *From Rational Choice to Reflexivity: Learning from Sen, Keynes, Hayek, Soros, and most of all, from Darwin*, *Economic Thought*, 3.1: 21-41, 2014.

White, William R., "Ultra easy monetary policy and the law of unintended consequences", *real-world economics review*, issue no. 63, 25 March 2013, pp. 19-56

Author contact: gustavoleomarques@hotmail.com

SUGGESTED CITATION:

Gustavo Marqués, "Processes vs. mechanisms and two kinds of rationality", *real-world economics review*, issue no. 68, 21 August 2014, pp. 10-24, <http://www.paecon.net/PAEReview/issue68/Marques68.pdf>

You may post and read comments on this paper at <http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

A systems and thermodynamics perspective on technology in the circular economy

Crelis F. Rammelt and Phillip Crisp [University of New South Wales, Australia / University of Amsterdam, The Netherlands and EcoSolve, Australia]

Copyright: Crelis F. Rammelt and Phillip Crisp, 2014

You may post comments on this paper at

<http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

Abstract

Several discourses on environment and sustainability are characterised by a strong confidence in the potential of technology to address, if not solve, the ecological impacts resulting from physically expanding systems of production and consumption. The optimism is further encouraged by leading environmental engineering concepts, including cradle-to-cradle and industrial ecology, as well as broader frameworks, such as natural capitalism and the circular economy. This paper explores the viability of their promise from a biophysical perspective, which is based on insights from system dynamics and thermodynamics. Such an ecological reality check is generally ignored or underestimated in the literature on aforementioned concepts and frameworks. The paper ultimately reflects on what role society can realistically assign to technology for resolving its ecological concerns. While environmental engineering undoubtedly has something to offer, it will end up chasing its tail if the social and economic forces driving up production and consumption are not addressed.

Keywords cradle-to-cradle, industrial ecology, circular economy, system dynamics, thermodynamics

1. Introduction

To manufacture complex infrastructures, products and services, engineering relies on inflows of natural resources from the planet's natural system in the form of energy and matter. The process also returns outflows of waste and emissions. Historical periods of economic activity have intensified these flows and their associated environmental impacts. Technology has played both aggravating and mitigating roles in the process.

Environmental engineering emerged in attempts to reduce the flows or their negative (side-) effects. Roughly until the first half of the 20th century, early environmental engineers assumed that the solution to pollution was dilution and dispersion. With the proliferation of industrial and consumer goods and the emergence of new forms of chemical waste, end-of-pipe solutions appeared after World War II, followed by pollution prevention strategies at the source. Engineers developed techniques for waste minimisation and recycling, as well as for improving resource efficiencies. These strategies were then integrated in the concept of cleaner production in the early 1990s. More recently, although largely based on older principles, several practices and concepts with strong engineering content claim to address the environmental impacts of industrial production and consumption without threatening economic expansion. These include cradle-to-cradle (McDonough and Braungart 2002; McDonough et al. 2003), industrial ecology (Frosch and Gallopoulos 1989; Graedel and Allenby 1995), natural capitalism (Hawken et al. 1999; Lovins et al. 1999) and the circular economy (Ellen MacArthur Foundation 2013). Other related concepts and designations include sustainable design, radical resource productivity, bio-mimicry, by-product synergy, technological food webs, industrial symbiosis and many more. Despite the changing terminology, the basic principles remain the same.

Common to these concepts is the belief that with the right innovations economic growth and environmental safekeeping can be complementary rather than in conflict. To explore the viability of this expectation, the global industrial machine must first be seen as a subset of a larger natural system, i.e. an ecological envelope (Boulding 1966). This envelope imposes several biophysical limits to human production and consumption. The combined insights from system dynamics (Meadows and Wright 2008) and thermodynamics (Corning 2002) will be used in this paper to examine some of these limits and to provide a reality check for aforementioned engineering concepts and frameworks. The paper's broader questions are about the role that we can realistically assign to engineering and about the point when we need to turn towards more fundamental social and economic transformations.

2. Understanding biophysical limits

The fields of system dynamics and thermodynamics provide valuable contributions to our understanding of biophysical limits, both quantitatively and qualitatively.

2.1 System dynamics

System dynamics tells us that a physically growing system dominated by reinforcing feedback¹ will eventually run into some kind of physical constraint, in the form of balancing feedback. In a growth-based industrial system, the more factories are operating, the more goods and services are produced and consumed. The resulting increase in profits leads to investments in new factories. Such a physically growing system relies on increasing inflows and outflows of energy and matter. In a bounded natural environment, whether balancing feedback originates from a non-renewable or a renewable resource makes a difference in how growth is likely to end, but not *whether* growth will end (Meadows et al. 1972; Turner 2008). We will now briefly explore both scenarios.

When an oil industry exploits a new oil field, profits are partly invested in establishing additional oilrigs, which leads to more oil extraction, higher profits and further investment in oilrigs. This represents reinforcing feedback. However, operations will first pick the proverbial low-hanging fruits. At some point, the extraction costs will outweigh the benefits and the resulting lower profits will reduce investments in new oilrigs. This represents balancing feedback. On the other hand, as oil becomes scarcer, prices go up and more money can be invested in new oilrigs, which pushes extraction upward. One feedback loop might dominate for a certain period of time and drive the system in a certain direction, but this doesn't mean that the other feedback loops have stopped existing. Meadows and Wright (2008) show that the potential lifetime of a newly discovered oil field available under the initial scale of operations is considerably reduced as a result of the dynamics at play.

A question to ask is: what if capital becomes more efficient instead of larger? Instead of expanding into new oil fields with additional oilrigs, more precise technology can be applied—for example through enhanced oil recovery techniques, such as gas or chemical injection. This can prolong extraction for a little while longer, but the upshot is a faster depletion of oil

¹ Systems can be understood in terms of stocks, flows and feedback. Stocks are accumulations of things (not necessarily physical) that change over time through the actions of inflows and outflows. Feedback occurs when changes in the size or composition of a stock affect the rates of inflow and/or outflow. Feedback is balancing or reinforcing, i.e. negative or positive.

towards the end (Meadows and Wright 2008). This indicates a fundamental role of technology in relation to ecology. It can act as a catalyst that speeds up the process of depletion of non-renewable resources.

We cannot engineer away the confines of a non-renewable stock of oil, coal, gas, iron, aluminium, copper, uranium or certain groundwater aquifers. What if we were to switch entirely to renewable natural resources?

Non-living renewables (sunlight, wind or rivers) regenerate through a steady input that keeps refilling the resource stock. Living renewables regenerate through reinforcing feedback: more fish means more reproduction and therefore more fish, for example. Another reinforcing feedback loop occurs when an increase in number of boats pushes up harvest, profits and investments in an even larger fishing fleet. At the same time, balancing feedback occurs as more harvesting means scarcer fish, which become more expensive to catch, reducing profits and lowering investments (Meadows and Wright 2008). Again, different feedback loops may dominate a system at different times.

In one situation, fish population and fleet size are in equilibrium, which can potentially maintain a steady harvest rate forever. However, a minor change can radically alter the outcome. Equivalent to the introduction of enhanced recovery technology in the oil industry, the introduction of bottom trawls or sonars maintains the yield per boat for just a bit longer despite dwindling fish populations. This can lead to overshoot and oscillations². With technology becoming even more efficient, the industry can wipe itself out entirely (Meadows and Wright 2008). This has been the fate of industrialised fishing in many parts of the world and there is evidence that we are now reaching global limits as well (Clover 2008). Again, technology functions as a catalyst that precipitates existing processes of growth and depletion. The dynamics may change with the introduction of fishing quota or other management systems, but this does not change the specific role of technology within the system, which is the focus of this paper.

Acknowledging the methodological limitations, the Global Footprint Network (2012) estimates it now takes the Earth one year and six months to regenerate what renewable matter we use in a year. In other words, stocks can act as buffers; the outflow from a stock of renewables can be temporarily higher than the inflow into that stock, but it will have to be compensated by lowering the outflow at some point in the future. So while non-renewable resources like oil are stock-limited, renewable resources like fish are flow-limited³. Similarly, if the rate at which we generate wastes exceeds the environment's ability to absorb them, this will have to be reversed in the future. For example, oceans and terrestrial ecosystems absorbed roughly 315 of a total of 555 gigatonne of accumulated anthropogenic carbon emissions (GtC) in the period 1750-2011. According to the IPCC, we have 50% chance of avoiding dangerous climate change if emissions stay below 840 GtC (Stocker et al. 2013). While emissions continue to grow, the absorption capacity of carbon reservoirs is limited and will eventually tail off (Ballantyne et al. 2012). There are biophysical limits to the amount of waste that can be

² Overfishing one year occurs at the expense of catches in the following year. Balancing feedback (fewer fishing boats) temporarily brings back fish populations, but overfishing reoccurs again the next season.

³ Strictly speaking, fossil fuels are also flow-limited if we can wait long enough for them to form. The process for currently exploited coal/oil/gas reserves is believed to be on the order of millions of years.

stored by the environment (the finite size of a sink⁴) and to the magnitude of the waste flows that can be absorbed and cycled over time (the renewable capacity of a sink) (Daly and Townsend 1993).

The simplified dynamics of the oil and fishing economies help us understand basic biophysical limits to the flows between engineered and natural systems in a quantitative sense. These flows also have certain qualitative characteristics that need to be taken into account. For this, we draw from the field of thermodynamics: a branch of physical sciences concerned with how energy changes from one form to another.

2.2 Thermodynamics

Thermodynamics states that energy can neither be created, nor destroyed; it can only change form. With a melting ice cube, heat transfers in one direction: from the surroundings to the ice. The heat lost by the surroundings equals the heat gained by the ice cube. The process will continue until there is equilibrium and the water has evaporated. Heat transfer from a colder system to a warmer environment can only be done by applying “work”. A fridge performs this work by taking heat out of the water in the ice cube tray inside the compartment and transferring it to the warmer kitchen. Thermodynamics also states that no energy transfer is 100% effective because of losses. In our example, it means that more heat is pumped into the kitchen than the amount of heat removed from the ice tray because of heat losses in the fridge’s electric wiring and from friction in the compressor.

Here, a distinction should be made between exergy and energy. Exergy represents the work potential, i.e., the useful portion of the energy used by the fridge to freeze water in the ice tray. While energy cannot be destroyed, exergy can. In other words, the fridge degrades some of the useful electricity into useless disorganised heat dissipated in the room. Energy is always tending toward more disorganised forms. The overall result is an increase in the degree of disorder or randomness, which is called entropy. We can see this in nature; everything perishes, rots, decays, falls apart and has the tendency to go from order to disorder⁵. In our example, the fridge and its components will eventually break down if we fail to apply work for their maintenance. The whole industrial system producing the fridge is bound to the same rule. Refineries transform crude oil into hydrocarbons and plastics; factories transform the hydrocarbons into kinetic energy, thermal energy and carbon dioxide emissions. Plastics degrade and end up in the environment in the form of micro- and nano-particles. This sequence of transformation increases entropy (decreases order).

The increase of entropy on earth as a whole is reversed only because of the existence of a complex biosphere powered primarily by solar radiation, which represents the main source of work and inflow of exergy. After most of this exergy is reflected back into space, some of it is transformed by plants and organisms into chemical exergy and some of it eventually ends up buried as low entropy stocks of carbon, coal, oil and gas. Flows of energy on earth are part of an open cycle; solar exergy comes in and heat goes out. Flows of matter on the other hand are part of a closed cycle (Boulding 1966). Ecosystems are driven by high-exergy and low-entropy resources, and generate almost no waste. In contrast, engineered systems are driven

⁴ A sink is a place in the environment where a compound or material collects. It can provide a natural pollution removal process or act as a reservoir that takes up a pollutant.

⁵ These changes are not only caused by entropic transformation, contributing factors include gravity, earth movement, wind, weather, solar radiation, oxidation and human use (Corning 2002).

by the extraction of low-exergy resources. At the other end, they produce, accumulate and dispose high-entropy emissions and waste (Nielsen 2007). This flow of energy and matter from ecological sources through the economy and back to ecological sinks has been referred to as “throughput” (Daly and Townsend 1993).

The distinction between natural and engineered systems does not mean that the former are purely frugal and cyclical, or that the latter are purely wasteful and linear. Many industries rely on the recycling of matter and energy from production processes and from consumption wastes. At the same time, the biosphere “dumps” carbon, coal, oil and gas in natural landfills (Jensen et al. 2011). While it is therefore wrong to set natural and engineered systems on opposite sides of a spectrum, there are nevertheless important differences. Nielsen and Müller (2009) argue that in natural systems, the cycles are local, decentralized and develop towards being increasingly closed with decreasing emissions and waste as a consequence. In engineered systems, however, the cycles are increasingly global, transport-intensive and have evolved to be open with increasing emissions and waste as a consequence. Waste control generally reduces profitability; costs therefore tend to be externalised.

3. An ecological reality check

For industrial systems, a low throughput of matter and energy implies a smaller ecological footprint and greater life expectancy and durability of goods and infrastructure; a high throughput implies more depletion of resources that will need to be renewed and more waste that will need to be disposed of (Meadows and Wright 2008). System dynamics and thermodynamics tell us that a tolerable rate of throughput and entropic transformation is ultimately dictated by the natural system, not by economics or engineering.

A possible task for engineering, within limits, would be to maximise the durability of stocks by minimising inflows of low entropy natural resources and by minimising outflows of high entropy waste and emissions. The role that industrial societies have assigned to technology is, however, much more Herculean. We have asked it to simultaneously and boundlessly minimise environmental impacts and maximise economic growth. In 1966, Kenneth Boulding suggested: “We are very far from having made the moral, political, and psychological adjustments which are implied in this transition from the illimitable plane to the closed sphere” (Boulding 1966: 2-3). How far are we now, almost half a century later?

3.1 Technology and substitution

Daly and Townsend (1993) see three dominant views in society. Some simply dismiss ultimate general scarcity on earth. Others accept the idea, but perceive the world as sufficiently large relative to the scale of human activity. Many have attempted to quantify the claim that engineered systems place on planetary resources (Vitousek et al. 1986; Haberl et al. 2007). There is tremendous uncertainty in the estimates, not in the least in determining the maximum scale that would lead to crisis levels (O’Neill 2011). A third dominant view in society sees human ingenuity and technical efficiency as the ultimate resources. There may be others, but the last position is the most interesting for the purposes of this paper.

While its rise to ascendancy is relatively recent in human history, capitalism in different shapes continues to spread as the “operating system” for most economies in the world. It

operates on the assumption that the output of production is a function of capital, labour and natural resources; shortages in the third factor elicit development of substitutes and higher efficiency in the first two factors. The suggestion in economic textbooks is generally that natural resources are not a limiting factor (Solow 1974; 1997). If we follow the logic of the production function we could ultimately bake a cake with only the cook and his kitchen; we do not need flour, eggs and sugar. We could also make our cake a thousand times bigger with no extra ingredients, if we stir faster and use bigger bowls and ovens (Georgescu-Roegen 1975; Daly 1997). In reality, of course, there are biophysical limits. The response to this is generally that markets will adjust production to impending environmental constraints (Solow 1997; Stiglitz 1997). While this may indeed occur at a local level, the real question is whether this will also occur at the aggregate scale of the global ecological envelope.

3.2 Decoupling, efficiency and effectiveness

Virtually all economies are currently growing both physically and financially, within a global envelope that is finite, non-growing and materially closed. A prevailing view, such as within the OECD and UNEP, is that the physical growth of throughput can be decoupled from the non-physical (financial) growth of GDP through innovation, which is commonly branded as “green growth” or “sustainable growth”. This view is also reflected, for example, in policy proposals for the next United Nations Climate Change Conference that emphasize decoupling emissions from growth (European Commission 2014). Two forms of decoupling are discussed in the literature: With relative decoupling, the growth of environmental impacts slows down relative to GDP due to efficiency improvements. With absolute decoupling, the environmental impact decreases as GDP grows (OECD 2002; Fischer-Kowalski and Swilling 2011).

To perpetuate a growing GDP under conditions of absolute biophysical limits will require—it is argued—compensation in terms of absolute decoupling of both the inflows from and the outflows into the environment⁶. Relative decoupling will not suffice; it will merely delay the point in time when one or more limits are reached (Blauwhof 2012). Moreover, absolute decoupling will have to be achieved on a global scale, because improvements in one part of the world might be achieved when production and associated ecological impacts are moved offshore (Bunker 1996; Bringezu et al. 2004).

There is evidence that global absolute decoupling has not occurred for important inflows of energy and matter. Global electricity consumption grew by 3% per year in the period 1980-2001. Reflecting improvements in energy efficiency as well as a shift towards less energy-intensive industries, global energy consumption still grew at a rate of 1.7% (EIA 2014). In the period 1980-2008, the amount of energy and raw materials required to produce a dollar of world GDP was reduced by 20%. At the same time, world GDP (in constant prices) grew by 125% such that total resource use still increased by 79% (SERI 2013). Only relative decoupling has been achieved (see Figure 1).

Turning now to the question of decoupling the outflows of waste and emissions, some have used carbon dioxide (CO₂) emissions as a proxy for other outflows. Global Real GDP (adjusted for inflation) grew by 3% per year in the period 1980-2001 and global CO₂ emissions grew 1.2% per year in the same period (EIA 2014). Between 1996 and 2006, these figures were 3.1% and 2.4%, respectively (Mitchell 2012). According to the IPCC, economic

⁶Strictly speaking, we should speak of “negative coupling” because GDP growth and throughput are still coupled; the coupling is simply functionally different (Smith and Max-Neef 2011).

growth was the main driver for global greenhouse gas emissions to grow more quickly between 2000 and 2010 than in each of the three previous decades (Edenhofer et al. 2014). Only relative decoupling has been achieved (see Figure 1). It is important to note that a trend in a subset of pollutants, such as CO₂, says little about the total environmental degeneration caused by a society. Some have suggested that energy use is a better approximation and this has never decreased in absolute terms anywhere with recorded GDP growth (Smith and Max-Neef 2011).

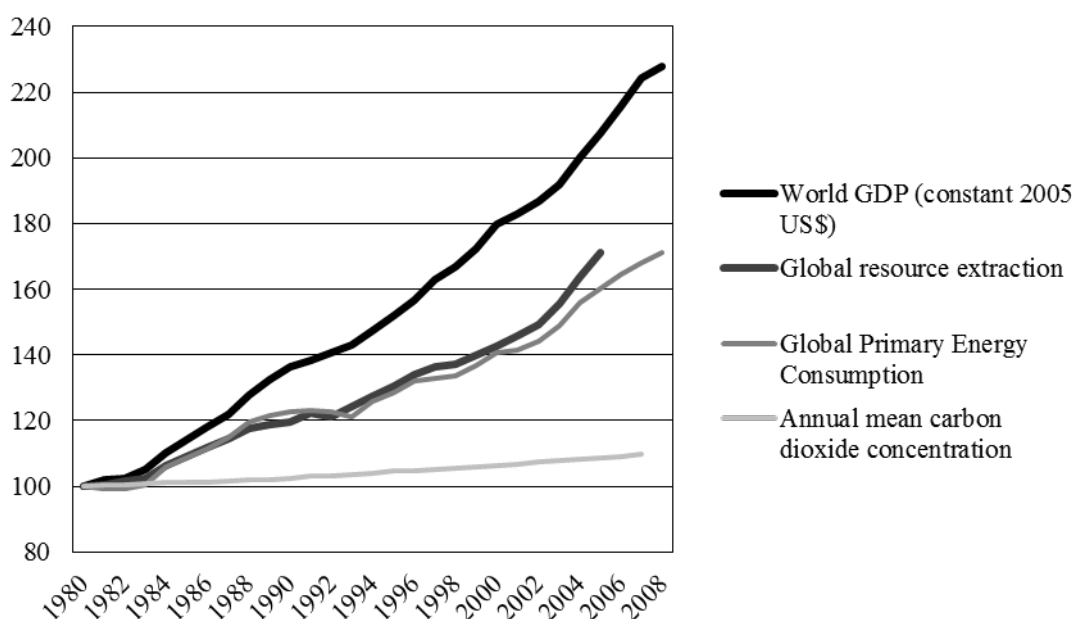


Figure 1 - Relative decoupling of growth from matter and energy flows.

Sources: World GDP, Resource extraction, Energy consumption, CO₂ concentrations based on World Bank (2014), SERI (2013), EIA (2014) and IPCC (2014), respectively.

We might be tempted to think that relative decoupling is only the first step towards absolute decoupling. However, relative decoupling is by no means a new phenomenon. Bunker (1996) describes how raw-materials-saving processes are older than the industrial revolution. Since the 16th century, innovation increased the strength per unit weight of metal, reduced the amount of copper required to transmit electricity, brought down the weight of charcoal needed to produce a ton of iron and so on. More efficient production processes replaced their more material-intensive predecessors, but they did not slow down the absolute growth of inflows of matter and energy (Bunker 1996). If relative decoupling indeed precedes absolute decoupling, the transition is seriously protracted.

There is also reason to believe that relative decoupling in the current economic system is making matters worse. Not only did the age-old dematerialisation strategies fail to neutralise overall growth of material production and consumption, they actually fuelled it. In economics, this process, known as the Jevons' paradox, was first described in 1865. The strategies contributed to reducing unit costs of production, which accelerated the circulation of capital, which in turn cheapened and intensified the appropriation of more natural resources. This represents reinforcing feedback whereby technology acts as a catalyst for increasing throughput. Dematerialisation has therefore usually been temporary, reflecting the lag

between the cost reduction and the expansion of production (Bunker 1996; Bringezu et al. 2004).

The question that comes up is whether the process of relative decoupling through efficiency improvements can make way for a process of absolute decoupling and ecological recovery through an entirely different set of engineering strategies. A distinction is therefore sometimes made between eco-efficiency and eco-effectiveness. The former improves by reducing the added environmental impact while maintaining or increasing the value of the output produced. The latter focuses on the development of products and industrial systems that maintain or enhance the quality and productivity of materials through subsequent life cycles (Braungart et al. 2007).

We will now explore how leading environmental engineering concepts and frameworks address either efficiency or effectiveness.

4. Environmental engineering and false expectations

From the above, it is clear that flows of matter and energy through the global economy have increased in absolute terms. Technological eco-efficiency has not been able to compensate for the expansion and may even have added fuel to the fire. Nevertheless, the mainstream sustainable development movement has trusted heavily in technology for solving the conflict between growth and the environment (WCED 1987; Weizsäcker et al. 1997; Schmidt-Bleek and Weaver 1998). This position is again very prominent in the eco-economic decoupling and green economy discourses (Brand 2012).

Several approaches with strong engineering content help perpetuate the promise:

1. The cradle-to-cradle framework “posits a new way of designing human systems to eliminate conflicts between economic growth and environmental health resulting from poor design and market structure” (McDonough et al. 2003: 436).
2. “Industrial Ecology is the means by which humanity can deliberately and rationally approach and maintain a desirable carrying capacity, given continued economic, cultural, and technological evolution” (Graedel and Allenby 1995: 9).
3. Natural capitalism incorporates “business strategies built around the radically more productive use of natural resources [that] can solve many environmental problems at a profit” (Lovins et al. 1999: 145).
4. The circular economy aims for a “transformation of products and their associated material flows such that they form a supportive relationship with ecological systems and future economic growth” (Ellen MacArthur Foundation 2013: 23).

There is overlap between the approaches, but their principles can be categorised as operating at different economic scales. Some are concerned with environmental pressures of the output of production, i.e., consumer products and services; others with environmental pressures at the production system level. The following paragraphs probe the expectation that these approaches will enable continued economic growth in an environmentally benign way in the long run. The conclusion is that such a view ignores, misinterprets or underestimates the biophysical limits outlined by the system dynamics and thermodynamics perspectives.

4.1 Overestimating approaches at the product level

Only “1% of all materials mobilised to serve America is actually made into products and still in use six months after sale” (Lovins et al. 1999: 152). Not only is there scope to redesign products in ways that reduce the squandering of material resources during manufacturing, there is also scope to reprocess much of the waste matter and (components of) the discarded products themselves.

Ecodesign, or “cradle-to-grave” design, seeks such improvements by considering the whole lifespan of a product⁷ (Brezet et al. 1997), but it has been criticised for placing the onus on consumers to dispose of products responsibly and for failing to address the physical limitations of the recycling process itself. While some materials like pure steel, aluminium, copper can be recycled indefinitely; others, such as paper, wood and plastics, can only make it through the process a limited number of times before they are disposed in landfill or incinerated. This can also happen with metal because of hard-to-separate impurities or because they are generally mixed into alloys. A typical soda can, for example, consists of two kinds of aluminium which are melted together during recycling, resulting in a weaker product (McDonough and Braungart 2002). At each cycle some of the matter is lost or degraded; recycling is really “down-cycling” (Kay 1994: 14), reflecting the process of increasing entropy. Cradle-to-cradle design therefore proposes closed-loop approaches where “waste equals food” (McDonough and Braungart 2002). It takes the view that zero-waste will never be realised because this would contradict the laws of thermodynamics. “The quantity of the emissions is not the problem, it is the quality of the outputs that must be addressed by making the emissions healthy” (Braungart et al. 2007: 6). The literature suggests that this type of eco-effectiveness can be achieved when products and their components are designed to consist of technical and biological “nutrients”. The former will permanently move as pure and valuable materials within closed-loop industrial cycles. The latter will easily re-enter the water or soil without releasing synthetic materials and toxins.

This proposed strategy is not without risk and uncertainty. First, the permanent movement of “technical nutrients” in closed cycles would violate the entropy law for most industrial materials as mentioned earlier (Reay et al. 2011). It is unclear whether those materials can all be phased out and replaced with appropriate materials, at a profit. Second, the manufacturing of “biological nutrients” depends on large quantities of plant materials. This will increase the scale of human appropriation of the stocks and flows of the natural system through agro-industrial production. This will likely aggravate the age-old impacts of agriculture on biodiversity, soil quality and water availability. It will also add a third rival in an already tense “food versus fuel” competition over agricultural resources. Meanwhile, increased waste and emissions consisting of biological nutrients would participate in biogeochemical cycles. An increase of inputs in those cycles can cause significant environmental damage, such as eutrophication from nutrient enrichment for example (Reijnders 2008).

One only needs to look around at what is on sale in shopping malls to see that cradle-to-cradle is much less widespread than another form of product engineering: design for obsolescence, which is defined as a deliberate strategy of making a product become rapidly out-dated or unserviceable in order to ensure continual sales. It represents a positive development from a narrow yet dominant commercial perspective. Philip Kotler, for example,

⁷ The lifespan of goods consists of raw materials acquisition, manufacturing, transportation, distribution and use to final recycling and disposal.

stated that this is “the working of the competitive and technological forces in a free society—forces that lead to ever-improving goods and services” (The Economist 2009). Were cradle-to-cradle to be taken up, it is likely that these commercial forces will generate products consisting of recyclable or biological materials with very short life spans. As suggested by a leading European carpet manufacturer: “cradle-to-cradle makes planned obsolescence good” (Sibley 2011). Such a view ignores earlier mentioned agricultural and biological concerns.

Others have proposed to respond to the problem of obsolescence by replacing disposable consumer goods with so-called product-services (Stahel and Reday 1976; Hawken et al. 1999; McDonough and Braungart 2002). “Services” in the sense used here focus on the utilisation and performance of goods, as opposed to the conventional definition of financial, health and education services. For example, Xerox sells reproduction services instead of photocopiers and Interface sells floor-covering services instead of carpets. The rationale is that it is in the interest of the manufacturer to avoid “leasing” products that quickly become defective. The idea has been around for a long time, but it hasn’t fundamentally altered patterns of consumption. In the current economic and cultural setting, such a system does not (yet) significantly compete with rental systems or private ownership (Reay et al. 2011). Whether it will is not a key issue in this paper. A more relevant concern is that product-services also rely on a biophysical basis for their production, use and replacement (Tukker et al. 2006). In a growth economy, product-services will also lead to growing throughput, which will also eventually hit some form of limit.

An inherent constraint of environmental product design strategies is that even if the individual impacts of a product were minimised, the increasing flow of total products sold and disposed would lead to a rise of the aggregate ecological cost. A few examples have already been discussed. Another limitation is that the strategies do not address the structural environmental challenges of current modes of production. Some have therefore sought to redesign entire industrial systems.

4.2 Overestimating approaches at the industrial level

In *Natural Capitalism*, Hawken et al. (1999) suggested that it is difficult to imagine the enormous potential for resource productivity, just as it was impossible 250 years ago to imagine the boost in labour productivity that lay ahead. Heat waste and discarded by-products are seen as evidence of profound inefficiencies. The authors claim that the U.S. economy is not even 10% as energy efficient as the laws of physics allow (Lovins et al. 1999). While they define efficiency in the engineering sense of doing more with less, measuring both factors in physical terms, they also suggest that this will save money. A relative decline in the volume of raw materials used per unit of GDP is assumed to lead to a process of absolute reduction in resource extraction and pollution (Hawken et al. 1999; Lovins et al. 1999). Advocates of the circular economy also expect that “the decoupling of growth from the demand for resources will slow current rates of natural capital erosion” (Ellen MacArthur Foundation 2013: 85). Both frameworks assume that relative decoupling leads to absolute decoupling.

The first objection to this assumption was discussed earlier. Historical evidence has shown that when you get more from less, you just take advantage of the slack (Bunker 1996). A second objection is related to the existence of a maximum efficiency limit. Perpetual financial growth within the confines of absolute biophysical limits is hypothetically only possible if efficiencies in the throughput keep perpetually rising faster than the rate of growth. As we

know, the second law of thermodynamics dictates that efficiency can never improve above 100% (Blauwhof 2012).

Beyond eco-efficiency strategies that merely lead to relative decoupling, the natural capitalism and circular economy frameworks suggest developing industrial-scale eco-effectiveness strategies that will lead to absolute decoupling. Lovins et al. (1999: 10) suggest that it is possible to eliminate waste “by redesigning industrial systems on biological lines.” Similarly, “the circular economy takes its insights from living systems” (Ellen MacArthur Foundation 2013: 26). They allude to the somewhat older concept of industrial ecology in which wastes from one industrial process can serve as the raw materials for another, thereby reducing environmental impacts (Frosch and Gallopoulos 1989) or even closing cycles of matter as occurring in natural ecosystems (Graedel and Allenby 1995). Half a century ago, Boulding (1966: 5) already argued that in a closed system “all outputs from consumption would constantly be recycled to become inputs for production, as for instance, nitrogen in the nitrogen cycle of the natural ecosystem”.

To be clear, despite the oldness of these ideas, industrial ecology does not yet exist in a strict sense. Most of the current examples consist of technical or operational modifications for reducing waste in individual firms. The inter-industry coordination that does exist today relies on cascading waste into feedstock. This is a practice that can reduce (or slow down the growth of) material throughput, but it does not close material cycles (O’Rourke et al. 1996). Following China’s vision of a circular economy, for example, there have been worthy efficiency improvements in the establishment of matter and energy exchanges within eco-industrial parks. However, resource consumption and waste generation continue to increase (Tian et al. 2014). It is difficult to imagine closing a system that imports such vast amounts of raw material inputs and exports over a third of its production output (as fraction of GDP in 2006) (Koopman et al. 2008). For now, the Chinese economy seems more “spiralling” than “circular”.

If it did occur in the future, a widespread adoption of industrial ecology principles would have to deal with the matter of entropy. As we increase recycling at the industrial scale, we diffuse and lose more and more matter at each cycle and we generate growing waste and emissions (Daly and Townsend 1993). Approaching closed material cycles would then require separating and reprocessing high entropy wastes to return and reuse them as low entropy resources (O’Rourke et al. 1996).

It is quite possible to re-concentrate diffused materials, but such a reduction of entropy has to be paid for by inputs of energy. An industrial-scale shift from virgin to reprocessed materials will produce shifts in energy use. On the one hand, producing a ton of steel plate from iron ore is almost four times more energy intensive than recycling steel (Daly and Townsend 1993). On the other hand, recycling chemicals, such as solvents from dilute industrial waste streams, may result in net energy costs (O’Rourke et al. 1996). Whichever way the balance would initially tilt, in the end, full-scale industrial ecology within a growth-based economy will demand growing energy inputs. Hopes are set on solar-powered electricity generation and its non-damaging bountiful source of exergy. However, such a system also requires a material basis for the construction of solar cells, the transportation and storage of electricity. Its growth will also lead to increasing waste heat. “In regard to the energy system there is, unfortunately, no escape from the grim Second Law of Thermodynamics” (Boulding 1966: 6).

Even with infinite sources of renewable energy, closed cycles remain difficult to imagine for complex materials such as pesticides, fertilisers, coatings, lubricants, adhesives, inks, brake pads or tyres. It is even harder to imagine for highly dissipative emissions resulting from the combustion of fossil fuels (O'Rourke et al. 1996). Industrial ecologists, natural capitalists and circular economists therefore argue that these materials can be phased out, also without threatening economic growth. The case that is brought up time and time again, perhaps for lack of alternative, is that of the cutback in chlorofluorocarbons, which simultaneously delivered windfall profits for business. However, this took place in very specific economic and political circumstances. For many reasons, this hasn't reoccurred on such a scale for other toxic and dissipative materials (Maxwell and Briscoe 1997).

5. Conclusions

This paper started by asking what role can realistically be assigned to engineering, and when we would need to look beyond technology towards economic and social changes.

Environmental engineering has so far failed to bring about the level of absolute decoupling that is required to sustain the current economic system. Present expectations of dematerialisation, recycling and loop-closing should be tempered by the fact that these engineering principles have been around for a very long time and that their environmental gains have been overwhelmed by economic growth. Several practices and concepts with strong engineering content nevertheless promise an absolute reduction in the environmental impacts of production and consumption systems in growth-based economies. For several reasons, this is a false promise.

Cradle-to-cradle overestimates the potential to close (growing) cycles of "technical nutrients". It also ignores or underestimates the impacts of a shift to "biological nutrients". Industrial ecology, natural capitalism and the circular economy framework overestimate the capacity to close (growing) matter cycles in production systems (particularly when dealing with toxic or dissipative matter). Their proposed shift from products to services ignores or underestimates the required physical basis. Their advocates also ignore or underestimate the fact that energy cannot be cycled and the consequences for energy inflows and heat waste outflows. In general, thermodynamic considerations are not receiving sufficient attention in the cradle-to-cradle and industrial ecology literature. These doubts are also pertinent to the natural capitalism and circular economy literature that relies heavily on cradle-to-cradle and industrial ecology principles.

Within a growth economy, the adoption of these engineering practices and concepts might slow down the growth of throughput. At best, this merely delays the time it takes to reach the boundaries of the biophysical envelope. At worst, the resource and energy savings generate profits that are reinvested in growth, which doesn't delay, but speeds up depletion and pollution. The field of system dynamics may help to mentally reconcile these seemingly conflicting dynamics. Different feedback loops might dominate and drive (parts of) the system in different directions at different times.

An appreciation of biophysical limits and thermodynamics should be much more prominent in the fields of economics and engineering. The insights tell us that there are limits imposed on the quantity of non-renewable resources, the pace of regeneration of renewables, how much

emissions nature can neutralise, how quickly wastes can be absorbed, how often materials can be recycled, and so on. Although not in the scope of this paper, this brings up important questions about social and economic equity. As we cannot increase the size of the pie indefinitely, there are ethical and political concerns about its persistent and worsening lopsided distribution (Rammelt and Boes 2013).

In conclusion, our economies must vastly be remodelled despite the engineering illusions that vindicate business as usual. "Clean coal" is an obviously deceitful example of this, but even our more genuine technical efforts cannot fully close material cycles and certainly cannot close energy cycles. Perhaps they do not need to. The natural system has the capacity to absorb a certain amount of our waste and pollutants. It also has the potential to generate a constant inflow of renewable resources. Within bounds, engineering could serve to maximise the durability of stocks by minimising throughput. The engineering concepts and frameworks discussed in this paper surely have something to offer in this regard, but they will end up chasing their tails if we do not address the social and economic forces driving up production and consumption. This expansion is instigated by the economy and catalysed by technology, but is eventually bound by ecology.

References

- Ballantyne, A.P., Alden, C.B., Miller, J.B., Tans, P.P., & White, J.W.C. 2012. Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years. *Nature* 488(7409): 70-72.
- Blauwhof, F.B. 2012. Overcoming accumulation: Is a capitalist steady-state economy possible? *Ecological Economics* 84(2012): 254-261.
- Boulding, K.E. 1966. The economics of the coming spaceship earth. In *Environmental quality in a growing economy*, ed. H. Jarrett, pp. 3-14. Baltimore, MD: John Hopkins University Press.
- Brand, U. 2012. Green economy the next oxymoron? No lessons learned from failures of implementing sustainable development. *GAIA-Ecological Perspectives for Science and Society* 21(1): 28-32.
- Brezet, H., C. Van Hemel, H. Böttcher, and R. Clarke. 1997. *Ecodesign: A promising approach to sustainable production and consumption*. Paris: UNEP.
- Bringezu, S., H. Schütz, S. Steger, and J. Baudisch. 2004. International comparison of resource use and its relation to economic growth. *Ecological Economics* 51(1): 97-124.
- Bunker, S.G. 1996. Raw material and the global economy: Oversights and distortions in industrial ecology. *Society and Natural Resources* 9(4): 419-429.
- Clover, C. 2008. *The end of the line*. Berkeley: University of California Press.
- Corning, P.A. 2002. Thermoeconomics: Beyond the second law. *Journal of Bioeconomics* 4(1): 57-88.
- Daly, H.E. 1997. Georgescu-Roegen versus Solow/Stiglitz. *Ecological Economics* 22(3): 261-266.
- Daly, H.E., and K.N. Townsend. 1993. *Valuing the earth*. Cambridge, MA: MIT Press.
- Edenhofer, O., R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Seyboth, ... and J.C. Minx. 2014. Climate Change 2014, Mitigation of Climate Change. Intergovernmental Panel on Climate Change, Working Group III Contribution to the IPCC Fifth Assessment Report. New York: Cambridge University Press.
- Ellen MacArthur Foundation. 2013. *Towards the circular economy vol. 2: Opportunities for the consumer goods sector*. Cowes, UK: Ellen MacArthur Foundation.

- Energy Information Administration (EIA). 2014. International Energy Statistics. <http://www.eia.gov/countries/data.cfm> (accessed 24 July 2014).
- European Commission 2014. Energy: Energy and climate goals for 2030. http://ec.europa.eu/energy/2030_en.htm (accessed 20 July 2014).
- Fischer-Kowalski, M., and M. Swilling. 2011. *Decoupling: Natural resource use and environmental impacts from economic growth*. Paris: UNEP.
- Frosch, R.A. and N.E. Gallopoulos. 1989. Strategies for manufacturing. *Scientific American* 261(3): 144-152.
- Georgescu-Roegen, N. 1975. Energy and economic myths. *Southern Economic Journal* 41(3): 347-381.
- Global Footprint Network (GFN). 2012. World footprint. http://www.footprintnetwork.org/en/index.php/GFN/page/world_footprint/ (accessed 12 April 2014).
- Graedel, T.E., and B.R. Allenby. 1995. *Industrial ecology*. New Jersey: Prentice Hall.
- Haberl, H., K.H. Erb, F. Krausmann, V. Gaube, A. Bondeau, ... and M. Fischer-Kowalski. 2007. Quantifying and mapping the human appropriation of net primary production in earth's terrestrial ecosystems. *Proceedings of the National Academy of Sciences* 104(31): 12942-12947.
- Hawken, P., A.B. Lovins, and L.H. Lovins. 1999. *Natural capitalism*. Boston: Little, Brown and Co.
- Intergovernmental Panel on Climate Change (IPCC). 2014. Annual mean carbon dioxide concentration measured at Mauna Loa. ftp://ftp.cmdl.noaa.gov/ccg/co2/trends/co2_mm_mlo.txt (accessed 24 July 2014).
- Jensen, P.D., L. Basson, and M. Leach. 2011. Reinterpreting industrial ecology. *Journal of Industrial Ecology* 15(5): 680-692.
- Kay, T. 1994. Salvo in Germany - Reiner Pilz. *Salvo* 23(11/09): 11-14.
- Koopman, R., Z. Wang, and S.J. Wei. 2008. How much of Chinese exports is really made in China? *NBER Working Paper*, 14109.
- Lovins, A.B., H.L. Lovins, and P. Hawken. 1999. A road map for natural capitalism. *Harvard Business Review* 77(3): 145-158.
- Maxwell, J., and F. Briscoe. 1997. There's money in the air: The CFC ban and Dupont's regulatory strategy. *Business Strategy and the Environment* 6(5): 276-286.
- McDonough, W., and M. Braungart. 2002. *Cradle to cradle: Remaking the way we make things*. New York: North Point Press.
- McDonough, W., M. Braungart, P.T. Anastas, and J.B. Zimmerman. 2003. Applying the principles of green engineering to cradle-to-cradle design. *Environmental Science and Technology* 37(23): 434-441.
- Meadows, D.H., and D. Wright. 2008. *Thinking in systems*. White River Junction, VT: Chelsea Green.
- Meadows, D.H., D.L. Meadows, J. Randers, and W.W. Behrens. 1972. *The limits to growth*. London: Earth Island.
- Mitchell, R.B. 2012. Technology is not enough climate change, population, affluence, and consumption. *Journal of Environment and Development* 21(1): 24-27.
- Nielsen, S.N. 2007. What has modern ecosystem theory to offer to cleaner production, industrial ecology and society? *Journal of Cleaner Production* 15(17): 1639-1653.
- Nielsen, S.N., and F. Müller. 2009. Understanding the functional principles of nature - proposing another type of ecosystem services. *Ecological Modelling* 220(16): 1913-1925.
- Organisation for Economic Co-operation and Development (OECD). 2002. Indicators to measure decoupling of environmental pressure from economic growth.

[http://search.oecd.org/officialdocuments/displaydocumentpdf/?doclanguage=en&cote=sg/sd\(2002\)1/final](http://search.oecd.org/officialdocuments/displaydocumentpdf/?doclanguage=en&cote=sg/sd(2002)1/final) (accessed 12 April 2014).

O'Rourke, D., L. Connelly, and C.P. Koshland. 1996. Industrial ecology: A critical review. *Environment and Pollution* 6(2/3): 89.

O'Neill, D.W. 2011. Measuring progress in the degrowth transition to a steady state economy. *Ecological Economics* 84(2012): 221-231.

Rammelt, C.F., and J. Boes. 2013. Galtung meets Daly: A framework for addressing inequity in ecological economics. *Ecological Economics* 93(2013): 269–277.

Reay, S.D., J.P. McCool, and A. Withell. 2011. Exploring the feasibility of cradle to cradle (product) design: Perspective from New Zealand scientists. *Journal of Sustainable Development* 4(1):36-44.

Reijnders, L. 2008. Are emissions or wastes consisting of biological nutrients good or healthy? *Journal of Cleaner Production* 16(10): 1138-1141.

Schmidt-Bleek, F., and P. Weaver. 1998. *Factor 10: Manifesto for a sustainable planet*. Austin, TX: Greenleaf.

Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, ... and P.M. Midgley. 2013. Climate change 2013: The physical science basis. Intergovernmental Panel on Climate Change, Working Group I Contribution to the IPCC Fifth Assessment Report. New York: Cambridge University Press.

Sustainable Europe Research Institute (SERI). 2013. Trends in global resource extraction, GDP and material intensity 1980-2008. <http://www.materialflows.net/home> (accessed 12 April 2014).

Sibley, A. 2011. Eco-effectiveness and the triple top line: Cradle to cradle. <http://www.sustain2green.com/p/eco-effectiveness-and-triple-top-line.htm> (accessed 12 April 2014).

Smith, P., and M. Max-Neef. 2011. *Economics unmasked: From power and greed to compassion and the common good*. Cambridge, UK: Green Books.

Solow, R.M. 1974. The economics of resources or the resources of economics. *The American Economic Review* 64(2): 1-14.

Solow, R.M. 1997. Georgescu-Roegen versus Solow/Stiglitz - reply. *Ecological Economics* 22(3): 267-268.

Stahel, W., and G. Reday. 1976. *Jobs for tomorrow, the potential for substituting manpower for energy*. Brussels: CEC.

Stiglitz, J.E. 1997. Georgescu-Roegen versus Solow/Stiglitz - reply. *Ecological Economics* 22(3): 269-270.

The Economist. 2009. Idea: Planned obsolescence. <http://www.economist.com/node/13354332> (accessed 12 April 2014).

Tian, J., W. Liu, B. Lai, X. Li, and L. Chen. 2014. Study of the performance of eco-industrial park development in China. *Journal of Cleaner Production* 64(0): 486-494.

Tukker, A., G. Huppes, J. Guinée, R. Heijungs, A. de Koning, ... and B. Jansen. 2006. *Environmental impact of products*. Sevilla: IPTS & ESTO

Turner, G.M. 2008. A comparison of the limits to growth with 30 years of reality. *Global Environmental Change* 18(3): 397-411.

Vitousek, P.M., P.R. Ehrlich, A.H. Ehrlich, and P.A. Matson. 1986. Human appropriation of the products of photosynthesis. *BioScience* 36(6): 368-373.

Weizsäcker, E.U.V., A.B. Lovins, and L.H. Lovins. 1997. *Factor four: Doubling wealth, halving resource use*. London: Earthscan.

World Bank. 2014. GDP (constant 2005 US\$). <http://data.worldbank.org/indicator/NY.GDP.MKTP.KD> (accessed 24 July 2014).

World Commission on Environment and Development (WCED). 1987. *Our common future*. Oxford: Oxford University Press.

Author contact: crelis.rammelt@unsw.edu.au and phillip@ecosolve.com.au

SUGGESTED CITATION:

Crelis F. Rammelt and Phillip Crisp, "A systems and thermodynamics perspective on technology in the circular economy", *real-world economics review*, issue no. 68, 21 August 2014, pp. 25-40, <http://www.paecon.net/PAEReview/issue68/RammeltCrisp68.pdf>

You may post and read comments on this paper at <http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

‘Right back where we started from’: from ‘the Classics’ to Keynes, and back again

Roy H. Grieve¹ [UK]

Copyright: Roy H. Grieve, 2014

You may post comments on this paper at

<http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

Abstract²

The purpose of this paper is to highlight the curiously circular course followed by mainstream macroeconomic thinking in recent times. Having broken from classical orthodoxy in the late 1930s via Keynes's *General Theory*, over the last three or four decades the mainstream conventional wisdom, regressing rather than progressing, has now come to embrace a conception of the working of the macroeconomy which is again of a classical, essentially pre-Keynesian, character. At the core of the analysis presented in the typical contemporary macro textbook is the (neo)classical model of the labour market, which represents employment as determined (given conditions of productivity) by the terms of labour supply. While it is allowed that changes in aggregate demand may temporarily affect output and employment, the contention is that in due course employment will automatically return to its ‘natural’ (full employment) level. Unemployment is therefore identified as a merely frictional or voluntary phenomenon: *involuntary* unemployment - in other words persisting demand-deficient unemployment - is entirely absent from the picture. Variations in aggregate demand are understood to have a lasting impact only on the price level, not on output and employment. This in effect amounts to a return to a Pigouvian conception such as targeted by Keynes in the *General Theory*. We take the view that this reversion to ideas which should by now be obsolete reflects not the discovery of logical or empirical deficiencies in the Keynes analysis, but results rather from doctrinaire blindness and failure of scholarship on account of which essential features of the Keynes theory have been overlooked or misrepresented. There is an urgent need for a critical appraisal of the current conventional macroeconomic wisdom.

Keywords Keynes's *General Theory*; ‘classical’ macroeconomics; involuntary unemployment; the AD/AS model

JEL Classifications B12, B22, E13

Introduction

The recent storm – indeed hurricane – which hit the world economy, bringing financial crisis, falling output and sharply rising unemployment along with the threat of general deflation, has called into question the validity of much contemporary macroeconomic theory. Conventional mainstream economics has in recent years been teaching that great macroeconomic disruption, such as experienced in the inter-war period, is a thing of the past – attributable to bungling mismanagement – and that, nowadays, thanks to more sophisticated theory and better economic management, we may safely relax in the confidence that modern advanced economies can be expected to operate at, or deviate only very temporarily from, their ‘natural’ – that is to say, full employment – level of activity. Optimism as to the benevolent working of market forces has been the keynote.³

¹ I am grateful to Eric Rahim for advice (if not always taken!).

² This paper derives from an article originally published in 2010 in the *GCU* (Government College University, Lahore) *Economic Journal*, XLI (2), pp.157-188, under the title ‘The Grand Macroeconomics Circular Tour’.

³ Not much attention seems to have been paid to the recent economic troubles of Japan.

When, however, in the autumn of 2008 the financial roof appeared to be falling in, and horrified commentators began to forecast for the world economy a re-run of the Great Depression of the 1930s – the name of J M (Lord) Keynes was increasingly mentioned. Keynes was suddenly remembered as the author of a powerful theoretical analysis which might again, as it did seventy years ago, provide understanding and guidance as to how to cope with conditions of collapsing activity across the world economy. Even five years into the present recession it is not clear that governments, with attention focused on debts, deficits and austerity programmes, understand how best to tackle the problem.⁴ For anyone presently seeking enlightenment on the nature of Keynes's ideas, it is not as easy as it ought be to find out what they are: the difficulty is that a different – older – interpretation of the working of the macro economy has come to the fore in contemporary analysis and discussion and that the Keynes vision has largely disappeared from mainstream teaching.⁵ We need to recover and distinguish the Keynesian vision from a conception which has come to obscure it.

With that end in mind, this paper sketches in broad terms the changing course of macroeconomic thinking (with reference to developed economies) on the causes of, and remedies for, unemployment. Beginning with the traditional 'classical' conception (which has shown a remarkable power of survival) we note how, from the late 1930s, that corpus of theory was supplanted by the significantly different Keynesian analysis. But the Keynesian conception, though for some time dominating the scene, has itself over the last thirty odd years been pushed into the background as old ideas of a pre-Keynesian, 'classical' sort (albeit in fashionable modern dress) have come back surreptitiously, but strongly, into vogue. We take the view that this rehabilitation of pre-Keynesian thinking represents a retrograde step: if contemporary economic problems are to be properly understood and handled effectively, it is, we believe, to the Keynesian tradition that a return must be made. This classical revival needs to be recognised for what it is, and the essential features of Keynes's analysis brought back out of the shadows into the light of day.

Accordingly in this paper we start by identifying the characteristic features of the pre-Keynesian theory, against which we set the essence of the significantly different conception advanced by Keynes in his *General Theory*. We then draw attention to the similarities of much contemporary mainstream macro theory to the old (pre-Keynes) classical analysis – a state of affairs which, as we interpret it, involves misunderstanding and misrepresentation of what Keynes was trying to say about the working of the macroeconomy.

Classical optimism with respect to aggregate demand: old style

In the early years of the nineteenth century, when the effects of technological and industrial change in boosting productive capacity were becoming evident, there arose amongst those with an interest in economic affairs a debate as to the possibility of a 'general glut' – a state of overproduction relative to demand for output across all industries within the economy. Was

⁴ Note the comment by Jeronim Capaldo (2013) 'Despite all contrary evidence, many institutions still recommend relying on fiscal austerity to stimulate growth.'

⁵ It is indicative of the unsatisfactory state of much standard economics teaching today that economics students at Manchester University are reported (*The Guardian*, 12/11/2013) to be fed up with the virtual monopoly of neoclassical doctrine, and are demanding the introduction to their courses of a wider spectrum of economic thought with greater relevance to real world conditions. *The Guardian* adds (21/11/2013) that 'the campaign is spreading fast: to Cambridge, Essex, the London School of Economics and a dozen other campuses and linking up with university groups in France, Germany, Slovenia and Chile'.

there a danger, it was asked, that the economy's ability to produce could come to exceed the willingness of the community to buy the product, thus giving rise to a problem of underutilised industrial capacity and unemployment?

A fierce controversy developed. On the one hand, proponents of what was to become the orthodox view (Say, James Mill, Ricardo, J S Mill) were confident that no problem of general excess supply could arise: they rejected out of hand the 'heretical' view (Malthus, Chalmers and Sismondi) that 'too much' investment might be undertaken, causing expansion of productive capacity to outrun the growth of demand. While it was recognized that the oversupply of any *individual* commodity could occur, a *general* state of overproduction – an autonomously-occurring deficiency of demand across the economy, was deemed an impossibility. Supply was said to 'create its own demand'.⁶ Advocates of this position cited 'Say's Law'- the proposition that the very act of supplying goods to the market implies a corresponding volume of demand – arguing that a producer was desirous either of consuming his own product or of exchanging it for the products of others. Essentially, therefore, the view was that desire to purchase could not fail to keep up with the volume of goods produced; even if savings were made out of income, such savings were not viewed as 'non-spending': income saved was expected to flow naturally to investment in capital goods. While it was admitted that monetary disturbances such as a reduction in the note issue, or an increased demand for cash in hand in a crisis, could give rise, at least temporarily, to unemployment, the orthodox view was that the ill-effects of such events would be transient; the basic Say's Law belief in the impossibility of a general glut on account of the production capacity of the economy exceeding the community's wants remained unshaken.

In the nineteenth century debates about the possibility of an overall deficiency of demand, it was the Say-Ricardo-Mill view that carried the day. As Keynes put it a hundred years later in a famous 'purple passage' (1936, p.32):

'The idea that we can safely forget the aggregate demand function is fundamental to the Ricardian economics, which underlie what we have been taught for more than a century, Malthus, indeed, had vehemently opposed Ricardo's doctrine that it was impossible for effective demand to be deficient: but vainly. For, since Malthus was unable to explain clearly (apart from an appeal to the facts of common observation) how and why effective demand could be deficient or excessive, he failed to furnish an alternative construction; and Ricardo conquered England as completely as the Holy Inquisition conquered Spain. Not only was his theory accepted by the city, by statesmen and by the academic world but controversy ceased. The great puzzle of "Effective Demand" with which Malthus had wrestled, vanished from economic literature . . . it could only live on furtively, below the surface, in the underworlds of Karl Marx, Silvio Gesell or Major Douglas.'

⁶ With hindsight we can say the proposition that 'supply creates its own demand' was erroneous for the reason that, while it is true that the act of production (supply) creates an equal amount of income (and thus purchasing power), planned spending by income recipients need not always be equal to the value of income earned or output which could be produced. In other words, while supply does create *ability* to buy (to purchase the output produced) it does not necessarily create at the same time an equal *will* to buy. Production and employment are adapted to the will to buy as expressed in the market – to the volume of effective demand.

Classical optimism with respect to aggregate demand: neoclassical style

Despite the fact that, from its emergence in the 1870s, neoclassical economics differed in certain other significant respects from classical political economy, with regard to the issue of aggregate demand, acceptance of Say's Law was, as Keynes implied in the passage just quoted, carried through into the new theoretical era, though given a characteristically neoclassical (marginalist) twist. Planned overall demand was still expected naturally to match supply. The rationalisation now brought forward focused on the so-called 'interest rate mechanism' to ensure the transformation of savings into intended investment.

In the neoclassical era, as had not been the case in classical times, the balancing of savings and investment at the full employment level of income was recognised as, at least in the short term, potentially problematical rather than simply automatic. It was held that establishment of the so-called 'natural' rate of interest ensured that all incomes generated through production were directly or indirectly returned as demand for output. Neoclassical writers did however allow that slow working of the interest rate mechanism could give rise to short-term deviations of employment and output from their full employment levels. That is to say, if a change was perceived in investment prospects, the *natural rate* would be expected to alter to maintain equality between savings and investment. There was though a possibility of trouble in that the banks could be slow in adjusting the *actual market* or *money rate* (that to which agents responded) to the change in the natural rate; if so, the consequence would be an excess or deficiency of intended investment above or below current savings. An excess of planned investment over savings or *vice versa* would then, via an increase or decrease of bank lending, cause, as the case might be, a rise or a fall in the money supply, that in turn implying an increase or decrease in spending.

What happened next with respect to employment and output, following an increase or decrease in spending, was held to depend on conditions of labour supply – specifically on the extent to which money wages responded to the change in prices (positive or negative) induced by the changed volume of monetary expenditure. If *money wages* responded immediately and fully to the change in prices, *real wages* would remain as before, and so correspondingly (the neoclassical theory held) would employment and output; the only effect of the disturbance would in these circumstances be a rise or fall of the price level. But if money wages failed to adjust immediately to match price changes (which was considered the more likely outcome), real wages would be affected, causing employment and output to alter. In other words, the neoclassical thesis was that stickiness of money wages in the face of price level changes meant that the real terms on which labour was being offered for employment were changing, with a direct effect on the quantity of labour taken into employment by employers.

In time, of course, once the money rate caught up with the natural rate of interest and real wages were restored to their 'proper' level, employment and output would return to their normal (full employment) levels. Cyclical unemployment as associated with such a sequence of events could be classified as 'frictional'.

It was however evident to Professor Pigou (a distinguished Cambridge authority) that the abnormally high and prolonged unemployment suffered in Britain in the inter-war period represented something other than the regular fluctuations of the trade cycle as had been experienced in earlier years. Pigou's diagnosis (propounded in his 1933 *Theory of Unemployment*) was that the distressing contemporary situation could be explained only as the result of an unduly high – permanently high – level of real wages having become established as normal. (See below, Figure 1.) He surmised that, after the dramatic changes in

prices and money wages that had occurred during the war and in the immediate post-war years, money wages had settled down in an inappropriate relationship to the level of commodity prices.

Pigou, that is to say, supposed that in the Depression years workers, in maintaining the going level of money – and so real-wages – were pricing themselves out of employment. The consequent unemployment could therefore be described as being, in effect, ‘voluntary’ – in the sense of being attributable to the decisions of the workers themselves. The remedy proposed was a cut in real wages. Pigou was confident that employment would then increase for the reason that with lower wages firms could be expected to move down their ‘labour demand (MPn) schedules’; he evidently took it for granted that the associated increase in output would be matched by a corresponding increase in real planned demand. The problem was not seen as one of want of demand. This is Say’s Law again. It was specifically on Pigou’s *Theory of Unemployment* that Keynes set his sights as providing the fullest and most explicit statement of what he described as the ‘classical’ position

We make no comment on the realism or otherwise of this neoclassical treatment of the labour market, other than to note its key feature - that *employment is said to be determined by factors wholly internal to the labour market*, i.e. by the (marginal) productivity of labour (Dn) and the conditions of labour supply (Sn). From this perspective, when wages fall, employment increases, *on the tacit presumption that demand for output increases sufficiently to take up the extra output produced with the higher level of employment*. Only if conditions in the product markets are supposed automatically to match whatever output corresponds to the going level of employment, can it be said that employment is determined simply at the point of intersection of Dn and Sn in the labour market.

Figure 1: The ‘classical’ (neoclassical) labour market

Figure 1(a): the standard model

Figure 1(b) Professor Pigou’s version

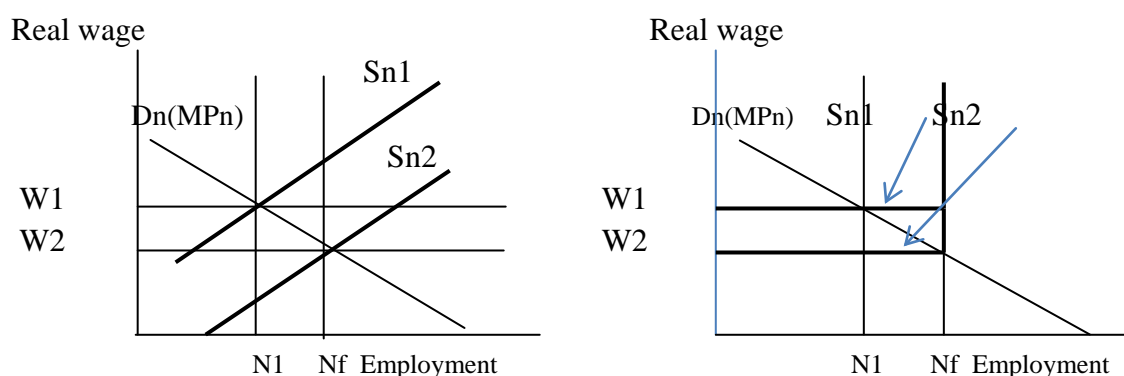


Figure 1 [1(a) and 1(b)] depicts the neoclassical representation of the labour market. The (so-called) labour demand curve Dn (MPn) is negatively-sloped, corresponding to the assumed diminishing marginal product of labour. In the more conventional representation [Figure 1(a)] the labour supply curve (Sn) slopes upward, implying increasing marginal disutility of work. In the Pigouvian version [Figure 1(b)], the labour supply curve is horizontal at the real wage for which the workforce ‘stipulates’.

Initially [in both Figure 1(a) and Figure 1(b)] with labour demand conditions (sic) indicated by D_n (MP_n), and conditions of labour supply as shown by S_{n1} , and with the real wage at W_1 , employment will be N_1 , which is less than N_f (full employment). To achieve full employment, all that is required (according to the neoclassical theory) is that the workforce accept a lower real wage, = W_2 . If they do, the labour supply curve shifts downwards to S_{n2} and employment will increase until the marginal product of labour again equals the value of the (now reduced) real wage. The establishment of this lower wage rate - workers have adopted more realistic assessment of what is an acceptable reward for getting up in the morning – is the necessary and sufficient condition for employment to increase from N_1 to N_f . No question is raised as to the validity of the assumption that demand can be relied upon to take up the increased output made possible by increased employment.

With hindsight we can now say that the fundamental error of the classical account of how employment is determined was the failure to integrate into the theoretical analysis the fact that *demand for labour is 'derived demand', labour demand depending on demand for the output that labour would produce*, depending that is to say, on conditions *outside* the labour market itself. In other words it was not understood that, even if labour supply conditions are fully compatible with full employment, labour could be out of work if demand for output in the product markets was not enough to justify employment of all the labour actually available for employment.

In summary, the pre-Keynesian orthodoxy, in both the 'old classical' period and the 'neoclassical' era, assumed that the real value of total spending would naturally match the available productive capacity of the economy. If unemployment existed, the cause lay not in an insufficient demand for the output of labour. Neoclassical theorists focused their attention on the *labour* market, not on the markets for output. With respect to both the short-term and the long, unemployment (as we have seen) was explained as being due to 'wrong' conditions of labour supply – with labour inadvertently or mistakenly asking for a rate of real wages incompatible with full employment. It is unfortunate that mainstream macroeconomic analysis today generally fails to recognise the implications of the Keynesian understanding that demand for labour is derived demand and continues to accept the neoclassical model of the working of the labour market.⁷

The Keynes Theory

By the mid-1930s Keynes had eventually arrived, he believed, at an understanding of what was wrong with the traditional analysis and of what was needed in its place. The short introductory chapter of the *General Theory* (1936, p.3) describes a radical ('revolutionary') agenda.

'I have called this book the General Theory of Employment, Interest and Money, placing the emphasis on the prefix general. The object of such a title is to contrast the character of my arguments and conclusions with those of the Classical theory of the subject, upon which I was brought up and which dominates the economic thought, both practical and theoretical, of the

⁷ As may be seen from a selection of standard macro texts; for instance: Abel, Bernanke and Croushore (2011), Burda and Wyplosz (2013), Froyen (2013), Gordon (2012), Krugman and Wells (2013), Mankiw (2013).

governing and academic classes of this generation, as it has for a hundred years past. I shall argue that the postulates of the Classical theory are applicable to a special case only and not to the general case, the situation which it assumes being a limiting point of the possible positions of equilibrium. Moreover, the characteristics of the special case assumed by the Classical theory happen not to be those of the economic society in which we actually live, with the result that its teaching is misleading and disastrous if we attempt to apply it to the facts of experience.'

As Keynes saw the situation, the classical theory failed to engage with the real-world conditions of the time – it failed to provide a believable explanation of the major contemporary (and world-wide) economic problem, that of high and persistent unemployment. The classical theory was, in his opinion, incapable of comprehending what had gone wrong:

'In addition to "frictional" unemployment, it (the Classical theory) is also compatible with "voluntary" unemployment due to the refusal of a unit of labour, as a result of legislation or of social practices or of combination for collective bargaining or of slow response to change or of mere human obstinacy, to accept a reward corresponding to the value of the product attributable to its marginal productivity. But these two categories of "frictional" and "voluntary" unemployment are comprehensive. The Classical postulates do not admit the possibility of the third category which I shall define as "involuntary" unemployment (Keynes, 1936, p.6).'

Keynes introduced the term '*involuntary* unemployment' to describe a situation in which workers, while perfectly willing to work on terms consistent with their being employed, were out of work because of a lack of jobs on offer. Such a state of affairs could emerge if employers, anticipating falling sales, cut back on output and employment, making workers, despite their not seeking any change in their existing terms of employment, redundant.

Keynes's explanation of the occurrence of involuntary unemployment depended on his identifying aggregate demand for output, *not* the conditions of labour supply, as the key *independent* determinant of production and employment within the economy. Aggregate demand was no longer treated as a 'tame' variable, ultimately tied to the value of output supplied. Keynes argued that if there happened to be insufficient demand within the system to justify the full employment of the workforce – as he believed was in fact the case in the early 1930s – workers would find themselves, against their wishes, without a job, but not on account of any action on their part in respect of wages. Involuntary unemployment occurs with want of demand relative to production capacity.

Keynes rejected both rationalisations previously offered for not worrying about the adequacy of aggregate demand. He saw aggregate demand as independent of supply, as an unstable and unreliable factor reflecting the planned expenditures of consumers and investors: there was no guarantee, as the old authorities such as Ricardo and J S Mill had maintained, that the very act of production implied a corresponding volume of planned demand. Neither, Keynes argued, could reliance be placed on the neoclassical notion of the 'interest rate mechanism': according to his new theory of liquidity preference, the role of the rate of interest was to reconcile asset preferences and supplies in the financial markets, and not to equate the flow of spending on new capital goods with the value of current saving.

This new conception of the working of the economy was expounded via a then novel macroeconomic model in which levels of output and employment depended on the total volume of demand, which was broken down into its component elements, the determination of each of which was analysed. Keynes's 'consumption function' postulated, for the first time, a key link between current income and the volume of consumption spending. With consumption (and savings) dependent on income, Keynes was able to explain how the economy responded to changes in demand through changes in output and employment – income changing until savings are brought into line with the current volume of investment spending. In dealing with (and he emphasised that this was crucial) a world of uncertainty, Keynes attached particular importance – because of its potential instability – to investment as a component of demand, stressing the dependence of investment on subjective factors of business confidence and expectations, factors liable to sudden and substantial revision. In times of particular uncertainty and perceived danger of loss, investors would avoid commitment to illiquid assets – such as new investment goods – preferring to keep their options open by reserving borrowing power or keeping unspent money in hand. (Nowadays consumer confidence and consumer demand may also be unstable.) With a collapse of confidence, the volume of effective demand would fall and the initial contraction would be amplified through the multiplier process.

In the Keynesian model of the income-expenditure (circular flow) system it was through changes in the level of activity that any imbalance arising between aggregate demand and the volume of current output was eliminated: output would (according to the circumstances) rise or fall, bringing savings (leakage from the circular flow of income and expenditure) into balance with investment (injection) until planned investment and savings were again equal. Changes in output and employment were now recognised as the natural, equilibrating response of the economy to changes in demand, not viewed as a temporary aberration. From the Keynesian perspective there was no reason to suppose the natural default position of the economy was that of full employment. Equilibrium is established when aggregate supply of output corresponds to aggregate demand, not necessarily when the demand for labour matches the supply of labour on offer. The system may get 'stuck' in a situation in which, although demand and supply of output produced are in balance, not all of the workforce are employed. In such a situation of *equilibrium with unemployment*, neither a reduction of money nor real wages would necessarily help. It was understood that the economy possessed no reliable adjustment mechanism by which, in the absence of a parametric change in the conditions of effective demand for output, the elimination of unemployment could be guaranteed.⁸

Demand deficient *involuntary* unemployment

Let us now focus our attention on Keynes's concept of *involuntary* unemployment.⁹ Keynes's identification of this hitherto unrecognised category of unemployment follows directly from

⁸ Keynes was of the opinion that, in recession, falling prices were more likely to worsen than improve the situation – a view with which non-neoclassical economists today would generally concur.

⁹ 'Involuntary unemployment' is a concept which has largely disappeared from contemporary mainstream macro texts. In none of the following texts - Abel, Bernanke and Croushore (2011), Blanchard (2011), Dornbusch, Fischer and Startz (2011), Froyen (2013), Gordon (2012), Krugman and Wells (2013) or Mankiw (2013) – does 'involuntary unemployment' appear in the index. Only Burda and Wyplosz (2013) refer to a phenomenon which they describe as 'involuntary unemployment', but to Burda and Wyplosz involuntary unemployment is occasioned by an excessively high level of real wages. They explain (p.118), by reference to the neoclassical labour market model: 'The solution to the

his rejection of the old, complacent view that aggregate demand could be relied upon (at least in due course) to match supply. As already mentioned, Keynes emphatically rejected the notion of aggregate demand being a 'tame' factor, as naturally corresponding to conditions of labour supply and output. What happens to output and employment reflects what is happening with the independent, potentially unstable, variable of aggregate demand. In explaining variations in employment, it is therefore necessary to look to conditions beyond the labour market – to the conditions prevailing in the product markets. Demand for labour, that is to say, is recognised as 'derived demand' – derived from the demand for output which justifies the employment of that labour in production. With a general lack of demand (relative to that required for full employment), the consequence of that deficiency emerges in the labour market as involuntary unemployment. As Keynes saw it, this is a misfortune that happens to workers as passive victims: demand for output can vary causing employment to rise or fall without the workforce having done anything to bring these changes about – or being able to do anything practical to remedy the situation. Demand-deficient unemployment is strictly 'involuntary'; in terms of the labour market diagram, labour is 'off its supply' curve. See figures 2(a) and 2(b).

Figure 2: Involuntary unemployment

Figure 2(a): Keynes 1936

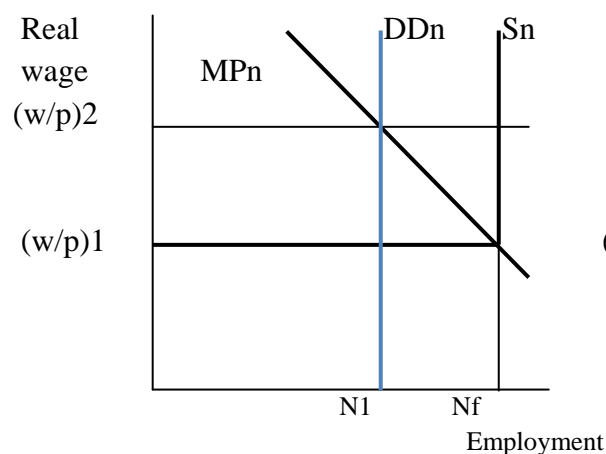
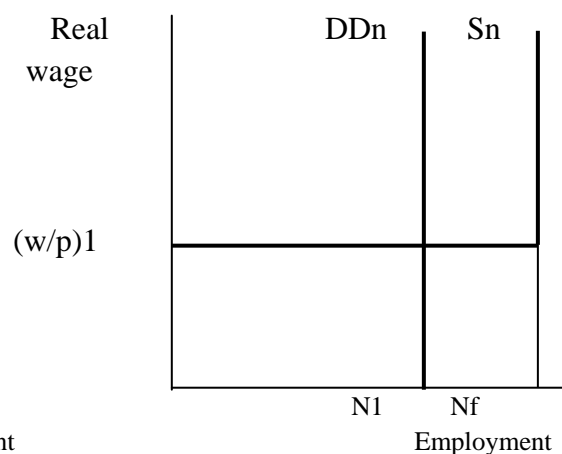


Figure 2(b): Keynes 1939



In figures 2(a) and 2(b) quantity of employment is measured on the horizontal axis and the real wage on the vertical axis. The labour supply curve (S_n) is drawn with a reverse L shape; this accords with the form suggested by Pigou, indicating that a given labour supply is available at a particular 'stipulated' rate of real wages. In figure 2(a) the downward-sloping marginal product of labour curve (MP_n) is shown. In neoclassical theory this is identified as the labour demand curve, but in Keynes's terms (bearing in mind that the demand for labour is derived demand) it shows not labour demand as a function of real wages, but an (inverse) relationship between employment (the independent variable) and wages (the dependent variable). The demand for labour is represented in the diagram by the vertical line DD_n

unemployment problem . . . appears rather straightforward. If the real wage were to fall from [its given value W_1 to W_2 , as in Figure 1(a)], demand would increase, supply would decrease, and unemployment would be eliminated. It is the failure of the [real] wage to decline that perpetuates unemployment. This is the key result: the existence of involuntary unemployment must be explained by **real wage rigidity**.' (Emphasis in original.)

(derived demand for labour), the position of which reflects conditions in the output markets and the quantity of labour required to produce the output demanded in these markets. In this model, employment is *not* determined at the point of intersection of the MP_n and S_n curves (as it is in the orthodox neoclassical theory) but *at the level of N at which the vertical DD_n curve intersects S_n*. DD_n will move left or right as demand for output falls or rises: full employment will be achieved only when the position of DD_n corresponds to N_f – that is to say, when demand for output is consistent with employment of all the labour available for employment. If, as in figures 2(a) and 2(b), the position of DD_n is not compatible with full employment, the gap in employment between N₁ and N_f corresponds to the quantity of demand-deficient *involuntary* unemployment.

Consider the sequence of events brought about by a fall in effective demand for output. Starting with a situation of full employment (DD_n corresponds to N_f), suppose that demand falls in the product markets, and DD_n shifts to the left as producers, unable to sell as much output as previously they could, lay-off workers. Involuntary unemployment equal to N_f – N₁ emerges.

There is a complicating quirk in this story as told by Keynes in the *General Theory* (1936) – a quirk subsequently eliminated (Keynes, 1939) – and, this is important – *eliminated with no change in the substance of the theory* between his 1936 and 1939 accounts. We need to be clear about this ‘quirk’. In the *General Theory* version, Keynes supposed that with demand for output declining and employers laying-off labour, firms would be moving leftwards down their positively-sloped short-run supply curves (corresponding to an increasing marginal product of labour/diminishing unit costs as output decreases), so that with costs and thus commodity prices falling, but with money wages unchanged, *real* wages would actually be rising. In terms of figure 1, that adjustment process is represented as a leftward movement up the MP_n curve; as production falls, employment declines from N_f to N₁, and at the same time (corresponding to the increase in the marginal product of labour with the lesser number of workers employed) the real wage rises from (w/p)₁ to (w/p)₂.

Thus we are presented in the *General Theory* with a scenario of *rising* real wages accompanying *falling* employment and output. This may look like the traditional classical theory, but Keynes was emphatic that it was not. The crucial difference between the Keynesian theory and the traditional theory is that in the Keynes theory, the increase in the real wage is no more than an incidental *consequence* of the fall in output and employment, *not the cause*; production has been reduced and with it employment, because demand for output has fallen, not because conditions of labour supply have autonomously altered. By contrast, the classical story is that the rise in the real wage is the initiating factor – the higher wage reduces the demand for labour and, as employment falls, so then does output; the rising real wage is the cause of the observed unemployment.

Keynes was, in 1936, under the impression that an inverse relationship between employment and real wages, such as he was assuming when composing the *General Theory*, was in accord not only with the standard (neoclassical) micro theory but also with the empirical evidence. However, soon after publication of the *General Theory* questions were asked as to whether real wages actually did move in the counter-cyclical manner predicted by the standard theory and assumed by Keynes so to do. New investigations revealed not only that the facts of the matter did not support the neoclassical prediction, but that Keynes, relying on Marshall’s report on the issue, had been misled by Marshall’s selective interpretation of the evidence available to him. It emerged that, over cyclical fluctuations, no regular inverse

relationship between employment and real wages in fact appeared to exist. That being so, Keynes (1939) realised that the actual course of events was much simpler than he had been supposing and that the complicated story told in the *General Theory* was redundant: when employment varied with changes in production, no associated changes in real wages required to be brought into the picture. Although Keynes's revised story was more straightforward than his earlier account, *the concept and explanation of involuntary unemployment remained exactly as before*. The diagrammatic treatment can however be simpler: the occurrence of demand-deficient involuntary unemployment may be demonstrated (as in Figure 1(b)) using a horizontal 'real wage–employment curve'¹⁰ in place of the downward sloping marginal product of labour schedule. When changes in demand for output cause changes in employment, real wages are seen to be unaffected.

Keynes's key insight, as revealed in the *General Theory*, had turned the traditional theory of employment on its head: as he explained the situation, the main line of causation linking the goods and labour markets ran not from the labour market to the goods markets, but *the other way*, from the goods markets to the labour market. Of great practical importance, the dominant economic problem of the day was now recognised to be one of *involuntary* unemployment, with its resolution to be found in stimulating aggregate demand, not in cutting wages.

After Keynes

In due course, indeed pretty quickly, the Keynesian theory became established as the new orthodoxy: a completely novel body of economic analysis – modern macroeconomic theory – developed. This was truly 'revolutionary'. Prominent in the new literature was the Hicks (1937) – Hansen (1953) IS/LM model which, integrating the income-expenditure and monetary elements of Keynes's system in a convenient diagram, was generally accepted as a satisfactory representation of the essentials of the Keynes conception. For more than thirty years, certainly until the late 1960s, Keynesian theory, though not unchallenged, formed the basis of mainstream macroeconomics. At the same time, in respect of practical policy, a new consensus developed. With a new understanding of the working – and possible malfunctioning – of the macro economy, it became generally accepted amongst academics, politicians, and the wider public that the government had an obligation to try to maintain an acceptable level of employment, and that it was feasible to use fiscal and monetary policy as means to that end.

The theoretical challenges to mainstream Keynesian orthodoxy that emerged in these first thirty odd years, although coming from sometimes hostile traditionalists, nevertheless implied acceptance of the essential Keynes proposition that aggregate demand was what mattered with respect to the determination of output and employment. The fact that attention was directed to the state of aggregate demand (even if different views existed as to the determination of aggregate demand) rather than, as in earlier times, to the level of real wages or to disparity between the 'natural' and 'money' rates of interest, indicated just how widely and profoundly thinking had been changed by the publication of the *General Theory*.

One critical line of argument explored in the 1950s and 1960s by theorists reluctant to accept the revolutionary Keynesian implication that the economy lacked any reliable 'self-righting'

¹⁰ Recognising that the real wage is invariant with respect to changes in employment.

capability, was built on the notion of a ‘wealth’ or ‘real balance’ (‘Pigou’) effect providing a possible rescue-mechanism for an economy sunk in heavy unemployment. The thesis was that a lower price level would mean a higher real value of the nominal money stock, thus generating a positive effect on spending.¹¹ If, the argument went, prices were to fall far enough, aggregate demand would be boosted to full employment level, regardless of any liquidity trap or interest inelasticity of investment demand. It was, however, explicitly recognised by theorists who investigated the potential of this mechanism (Pigou, Patinkin, and Keynes himself – who had in the *General Theory* given careful attention to the possible consequences of deflation) that the weakness of the wealth-effect on consumption, and even more importantly, the negative effects of falling prices on demand – rising indebtedness, expectations of further deflation – ruled the real balance effect out of court as a practical equilibrating mechanism. (Endogeneity of the money supply is a further nail in the coffin of the real-balance effect.) It was agreed that a decline of money wages and prices in a depression might actually make things worse rather than better. Patinkin’s summing up on the issue is worth noting (Patinkin, 1959):

‘The automatic adjustment process of the market is too unreliable to serve as the practical basis of a full-employment policy. In other words, though the real balance effect must be taken into account in our theoretical analysis, it is too weak – and, in some cases, too perverse – to fulfil a significant role in our policy considerations.’

Further reason to be sceptical of the argument that a process of deflation can have a positive effect on demand and employment is provided by recent experience in Japan, where falling prices have certainly not rescued the economy from recession. Thus J H Makin (2006), an informed observer of the Japanese scene, writes:

‘Deflation is dangerous. The nightmare of a deflationary spiral arises from the fact that as deflation intensifies and prices fall more rapidly, the real cost of borrowing rises. With a zero interest rate and 1 per cent deflation, the real cost of borrowing is 1 per cent. If deflation intensifies to 2 per cent, while the demand to hold cash strengthens because the rise in deflation represents a rising, risk-free, tax-free return on cash, more cash will be demanded. The move into cash further depresses spending, and thereby further intensifies deflation. The real cost of borrowing keeps rising, imparting an accelerating drag on the economy. . . . As noted, a deflationary spiral produces a sharp increase in the demand for liquidity that, if not satisfied by the central bank, will be satisfied by households and businesses selling goods and services, thereby intensifying the deflationary spiral.’

Far from welcoming falling prices as a means to recovery, the Japanese authorities were desperate to escape from deflation. So much for the practical relevance of the real balance effect.

Nevertheless, in contemporary textbook analyses it is virtually universal to include (as for instance in the form of the negatively-sloped AD curve of the AD/AS model) an apparently reliable inverse relationship between the level of prices and the volume of aggregate demand.

¹¹ The idea is that if the real value of people’s money holdings rise, they will feel better-off and consumption spending will increase.

While practical policy-makers fear deflation, textbook authors propound the proposition that falling prices should be welcomed as effective in lifting an economy out of recession.

The AD/AS model

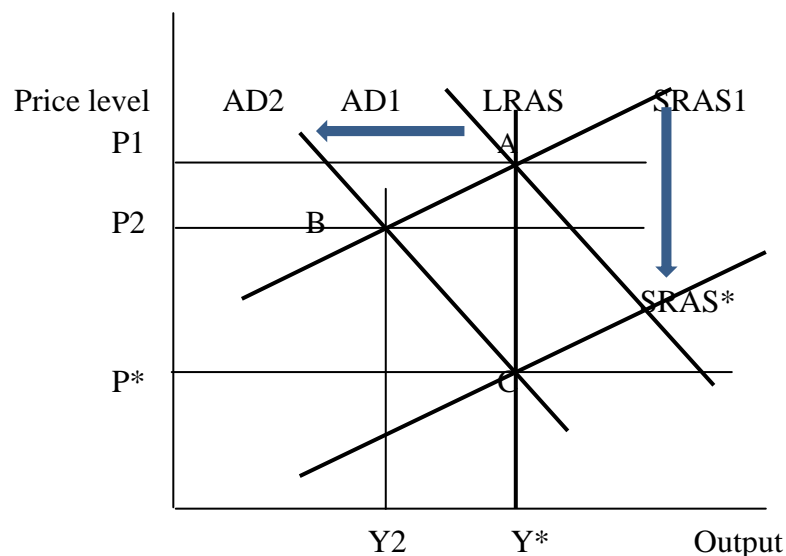
From some date around the 1970s the macroeconomic theory presented in the mainstream textbooks began to undergo a significant change. The principal contemporary challenge to the Keynes theory comes from the recent resuscitation of employment theory of a pre-Keynesian sort. The adoption of AD/AS as the expository model of choice for analysis of simultaneous changes in output and the price level has provided a vehicle by which a pre-Keynesian conception of the working of the macro system has been smuggled into, and established within, the mainstream of textbook teaching. Utilisation of AD/AS has not only been associated with a powerful resurgence of pre-Keynesian modes of thinking, it has had the further effect of obscuring and confusing the issue as to the nature of the fundamental difference of vision which separates the Keynesian and classical conceptions.

Recall how the AD/AS model comes into the story. In the typical textbook treatment,¹² when the exposition moves beyond the fix-price world of the Keynesian cross and IS/LM, the familiar story is that, within the economy, the volume of output (and with it employment) and the level of prices are simultaneously determined at the point of intersection of the so-called AD and AS schedules. (See Figure 3.) The negatively-sloped AD curve, derived from IS/LM equilibrium at different price levels, is said to represent the value of effective demand as a function of prices. The positively-sloped short-run AS curve shows output also as a function of prices; output changes occurring via alterations of real wages and labour supply in response to changes in spending. It is supposed that in time these inappropriate changes in real wages are corrected and the long-run vertical AS curve comes into play. Thus variations in effective demand for output are predicted to have no more than a temporary impact on output and employment but a permanent effect on the price level.

There is a fundamental issue respecting the internal consistency of this model. The fact is that the two halves of the model derive from incompatible theories of output and employment: they simultaneously represent different complete (and conflicting) accounts of the relationship between the price level and the level of output. Far from it being the case that the two schedules together (supposedly representing aggregate *demand* and *supply*) determine a unique equilibrium, each in itself represents a self-contained explanation of the equilibrium level of *output* (Y). If we focus on either curve, the other is redundant.

¹² The 'typical textbook treatment' may be found in, for example, Gordon (2012), Froyen (2013), Mankiw (2013) and Abel, Bernanke and Croushore (2011). Criticism of the AD/AS model on the grounds of internal inconsistency has been advanced by, amongst others, Rao (1991, 2007), Colander (1995), Grieve (1998, 2010) and Moseley (2010). Significantly also, from a 'New Classical' perspective, Barro and Grilli (1994) reject the conventional AD/AS model. Attempts have been made by Kennedy (1998) and Scarth (2010) to defend the model. It should however be noted that most, but not all textbooks, use the 'standard' AD/AS model with the deficiencies of which we are here concerned; Blanchard (2011) and Dornbusch, Fisher and Startz (2011), while keeping the usual label of 'AD/AS', in fact employ a quite different and self-consistent construction which does *not* make use of the usual classical-type AS curve (but which relies heavily on the highly suspect wealth/real-balance effect as a mechanism of adjustment).

Figure 3: The AD/AS model



With respect to the AD curve, both Rao (1991) and Colander (1995) have observed that, while it is described or defined as being analogous to a micro demand curve, it is in fact, as derived from IS/LM, not really a demand curve at all. Since what it shows in relation to levels of price are levels of income corresponding to IS/LM equilibrium, the AD curve is actually a locus of points of equality of aggregate demand and aggregate supply. Colander (1995) suggests the designation 'aggregate equilibrium curve'.

An equivalent situation exists with respect to the supposed supply side. The AS curve is evidently based on a pre-Keynesian understanding of the functioning of the labour market. The significant feature of that pre-Keynesian analysis is (as we have already emphasised) that the labour market is treated in exactly the same terms as the market for any final commodity – say, mangoes – ignoring the special nature of the market for labour services: labour demand and supply are both taken to be functions of the real wage. Changes in employment must then be attributed to movement of one or other curve: in the short run, with the marginal product of labour ('labour demand curve') fixed in position in accordance with the given technological conditions, it is only through movement of the labour supply curve that employment can alter. If employment is to change whenever (*ceteris paribus*) the labour supply curve shifts, and labour demand is to remain equal to labour supply, demand for output must be whatever is required to take up the output corresponding to the volume of employment determined in the labour market. In other words that neoclassical representation of the labour market depends for its validity on the assumption that aggregate demand for output is 'tame', i.e. that supply really does create its own demand. An aggregate supply curve based on these pre-Keynesian foundations therefore depicts, as a function of the price level, quantities of output which are matched by an equal volume of effective demand. That is to say, with aggregate supply matched by aggregate demand at all points along the curve, we have a second 'aggregate equilibrium curve'.

The standard AD/AS construction must be recognised as fatally flawed – its two component elements are fundamentally incompatible. Nevertheless textbook authors continue to make use of it. To be able to do so they must, wittingly or unwittingly, employ some strategy to obscure the unavoidable incoherence of any analysis derived from use of AD/AS. Two

complementary tactics can be identified: one is to treat the AD and AS curves as if they actually were (which of course they are not) macro equivalents of micro demand and supply curves; the other is effectively to neutralise one of the two conflicting theories of output and employment embodied in the AD/AS construction. It is the latter that particularly concerns us here: we find that the usage typically favoured in the textbooks essentially washes the Keynesian element out of the story, so that we are left with an analysis of a basically pre-Keynesian character.

Consider the standard textbook account. (See again Figure 3.) To make the model usable – although the reader is not warned of the trick being pulled – the AD and AS curves are treated as corresponding to ordinary micro demand and supply functions. Starting from an equilibrium position at A (Y^* , P_1) corresponding to employment at the ‘natural’ rate, a disturbance, for instance a sharp contraction of the money supply, occurs:¹³ the AD curve therefore shifts to the left, giving a temporary equilibrium with lower output and employment and some reduction of prices. The interpretation is that while spending and prices fall, money wages – either because of misperceptions or institutionally-determined stickiness – fail to keep pace with the falling prices, and as real wages accordingly rise, employment correspondingly falls. The economy has moved down the short run AS curve to position B (Y_2 , P_2). Subsequently, over time, money wages do adjust and the price level falls further as the short run AS curve slides down AD, so that the economy eventually reaches position C (Y^* , P^*). The system has then returned to its ‘natural’ (full employment) level of activity.

The key to understanding this supposed adjustment process is to appreciate that from the neoclassical perspective employment is – as we have continually emphasised – determined in the labour market at the point of intersection of the ‘labour demand’ (MP_n) and labour supply schedules, and that if the latter is temporarily displaced from its normal position (as it may be due to misperceptions about the existing relationship between prices and money wages) such errors will, it is confidently expected,¹⁴ readily be corrected, and conditions of labour supply will return to normal; and so, of course, will employment. *Employment varies from its ‘natural’ level because of an issue with the conditions of labour supply and the real wage; the natural level of employment is re-established when that issue is resolved. The position of the labour supply curve is all-important, and the presumption is that, following a disturbance, it will soon recover its ‘proper’ position.* As real aggregate demand for output *must* in equilibrium correspond to aggregate supply of output as determined by the conditions of labour supply, it follows that the price level *must* necessarily fall into line as required to ensure that the real value of spending corresponds to the output (indicated by the position of the vertical LRAS curve) associated with full employment in the labour market.

The stages of the disturbance and the subsequent adjustment process are described in terms of the temporarily changing conditions of labour supply – with real wages varying (whether on account of misperceptions or institutionally-determined wage stickiness). In the first instance, with reduced spending on output commodity prices fall and, though money wages may also fall (but not by so much), real wages actually rise, causing employment to decrease. At this point real wages exceed their equilibrium value (in real terms the S_n curve has risen). The neoclassical presumption is that the workforce quickly come to appreciate that (for whatever reason) wages have not adjusted *pari passu* with prices, and that that accounts for the

¹³ Typically for the textbooks, a non-Keynesian disturbance.

¹⁴ Recall Milton Friedman’s (borrowed) aphorism (Friedman, 1975, p.21): ‘... you can fool all of the people some of the time, you can fool some of the people all of the time, but you can’t fool all of the people all of the time.’

unemployment which has developed. Given current prices, the going level of wages represents a costly mistake. Further wage reductions are therefore accepted. As money wages fall, so do prices. So long as real wages and employment have not returned to their original ('natural') levels, money wages will continue to fall relative to prices, the labour supply curve will continue to shift downwards and employment will continue to increase. As the terms of labour supply to return to normal (the S_n curve having resumed its original position) employment and output recover to their original levels. *What is driving the labour market back to its previous equilibrium state is the acceptance on the part of the workforce that wage adjustments must be made for the restoration of that normal situation.* With real wages altered as required, employment returns to the natural rate. During this process of real wage adjustment, the general price level is also falling, and will continue to change until the real value of monetary spending matches the value of output produced when employment is at the natural rate.¹⁵

Given the supposed capability of the workforce to establish (if that is what they wish) a real wage consistent with full employment, and given the presumption that real aggregate demand can be relied upon to match full employment output, equilibrium can only be at the natural rate of employment, at which position (given current nominal expenditure) no pressures operate to alter money wages, real wages or prices. We are thus back to the conventional pre-Keynes (neoclassical) understanding of the working of the system: while it is allowed that changes in aggregate expenditure can, because of frictions, have a temporary effect on output and employment, in the longer term, the effects of such changes are solely on the price level.

What has happened to the Keynesian theory of demand? *There is evidently no place here for a theory which explains the equilibrium level of employment as determined by a factor independent of labour supply conditions – that is to say, by real effective demand for output.* In this analysis equilibrium employment depends simply by factors *internal* to the labour market – on conditions of labour supply against marginal productivity. While the AD curve is derived from Keynesian foundations (IS/LM), as it is used in this context its Keynesian ancestry is effectively neutralised. AD has no role as regards the equilibrium values of employment and real output which are determined – given the inclusion in the model of the neoclassical labour market and its associated Say's Law implication - within the labour market. The only role left for that shadowy remnant of the Keynes theory – the AD curve – is to set the price level at which the real value of monetary expenditure accords with the output which corresponds to the labour market equilibrium.¹⁶ In other words, use of the fashionable AD/AS model has permitted the development of an analysis which, although ostensibly building on Keynesian foundations, actually arrives at a conception which owes nothing to these Keynesian elements, but reproduces instead the conventional conclusions of pre-Keynesian orthodoxy.

¹⁵ To repeat for emphasis: the nature of the system is such that if the real value of nominal spending exceeds or falls short of full-employment output, the adjustment process just described, with money wages, real wages and prices all changing (rising or falling as the case may be) will continue until the real value of total spending corresponds to full-employment output.

¹⁶ We might as well have here, instead of the AD curve, the rectangular hyperbola of the quantity theory analysis which indicates the equilibrium price level, given real output and the money supply. While the Keynes theory explains the amount of *real* planned demand, in the standard AD/AS model, the AD curve is in effect treated as representing quantities of *nominal* expenditure, the real value of which can readily be rendered consistent with the real value of output as determined by conditions of labour supply.

Having thus been brought to a traditional 'classical' understanding of the determination of employment and output, the textbook reader may then as just well forget about the notion of planned real demand being something of significance. With the Keynes theory 'crowded out' of the story, out goes the concept of demand-deficient involuntary unemployment; whatever unemployment is envisaged from this perspective can only be frictional (presumed in time to be self-correcting) or voluntary.

'Right back where we started from'

We have indeed come round in a circle. The whole vision of the working of the macrosystem presented, in terms of the AD/AS model, by far too many contemporary textbooks, is essentially pre-Keynesian. Monetary spending may fluctuate, but whether or not such fluctuations affect employment and output is said to depend on reactions affecting real wages. Slow adjustment of money wages to price changes is held to account for cyclical variations in employment and output. With respect to the longer term, it is presumed that real wages return to their proper full-employment level. There are then no obstacles on the side of demand to prevent re-establishment of the 'natural' (full employment) level of activity. The pale shadow of Keynesian theory in the ADAS model – the AD curve – has nothing to do with the values of output and employment at equilibrium, only with the price level.

Finally, and further to emphasise the 'classical' character of this contemporary analysis, notice how closely the modern textbook conception corresponds to the views presented eighty years ago by Professor Pigou, whose *Theory of Unemployment*, was identified by Keynes as 'the fullest presentment' of the classical theory. As regards the short run, we find in Pigou (1933, pp.293-297) an account of the implications of wage stickiness, *virtually identical to the story told in present day textbooks*.

He refers to 'factors of inertia' operating on both sides of the labour market: these make employers reluctant to raise wage rates when conditions improve, and employees resistant to wage cuts when activity is declining:

'Thus, except in periods of very violent price oscillations, employers in general fight strongly against upward movements in money rates of wages and workpeople themselves against downward movements. Money wage-rates show themselves highly resistant to change.'

Pigou continues:

'These factors of inertia, which in an economy where wage-rates were always contracted for in kind, would tend to keep real wages stable in the face of changing demand, in a money economy tend to keep money wages stable ... In general, the translation of inertia from real-rates to money-rates causes real-rates to move in a manner not compensatory, but complementary, to movements in the real demand function. Real wage-rates not merely fail to fall when the real demand for labour is falling, but actually rise; and in like manner, when the real demand for labour is expanding, real wage-rates fall.'

It is to be understood – in accordance with the classical belief that employment is determined in the labour market at the point of intersection of the marginal product of labour curve and

the labour supply curve – that fluctuations in employment and output accompany these variations in real wages.

As regards the longer term, the Pigou theory is, again, essentially the same as that of the textbooks – that the equilibrium level of employment naturally tends to accord with the conditions of labour supply. The only difference is that Pigou saw wage rigidity as more of a problem – setting an unduly low limit to the equilibrium level of employment – than contemporary textbook authors appear to do. Although Pigou looked to inappropriate wage demands by the workforce and wage rigidity as preventing the establishment of equilibrium at a proper full-employment level, he had no doubts though that, whatever the situation in the labour market, aggregate demand for output could be relied upon to take up whatever volume of output corresponded to the existing conditions of labour supply and employment. Pigou (1933, p.73) expressed this ‘Say’s Law’ understanding in the following words:

If the real wage rate is reduced in the wage-goods industries, a powerful reaction is set up for an expansion in the demand for labour in the non-wage-goods industries ... it is certain that in practice the reaction on (employment) will be substantial.

The typical present-day textbook account differs from Pigou’s in assuming greater wage and price flexibility. Nevertheless, both Pigou and the modern authors share the vision that, if real wages can be kept at, or adjusted to, the appropriate level, there need be no anxiety about employment: we can then be confident that the real value of total expenditure will be sufficient to absorb whatever volume of real output is produced when the economy is operating at full employment. Pigou – tacitly – held by Say’s Law. The present-day textbook authors’ confidence that the economy tends to full employment stems essentially from their reliance on the neoclassical model of the labour market (which for its applicability requires the validity of Say’s Law); this indirect reliance on Say’s Law is however disguised by the inclusion in the analysis of an AD curve of Keynesian origin. But that is no more than an empty gesture in the direction of the ‘Keynesian revolution’: as we have noted, the AD curve plays no part in this explanation of the determination of the equilibrium level of employment and output.¹⁷

Thus as regards the fundamental elements of the Keynes conception – that planned aggregate demand is an independent and unreliable determining factor of output and employment, that deficiency of aggregate demand causes involuntary unemployment and that there is no reason to suppose that the ‘default’ state of the macroeconomic system is one of full employment – all have disappeared. The *General Theory* might as well never have been written. We are back in a ‘classical’ world in which the economy is understood to possess a natural tendency, when disturbed, to revert readily to full employment. But in the real world, things are unfortunately very different – the macroeconomy insists on misbehaving.

How have we got into this situation? In the 1970s, reflecting a general change in the political and intellectual climate, economic theorists and commentators of a right-wing, free-market

¹⁷ As mentioned above (footnote 8) certain contemporary macro texts (a very small minority) do not make use of the conventional ADAS model with the AS curve derived from the neoclassical labour market analysis. (In these texts, AS shows price as a function of output, not output as a function of price.) But the trouble with that version of ADAS is that the AD function (reflecting the real-balance effect – long past its sell-by date) plays a key part in ensuring adjustment to full employment equilibrium. Far too much weight is thus placed on a very questionable adjustment mechanism. Again the prediction of automatic adjustment to full employment is a far cry from Keynes’s view.

persuasion began to advance, with renewed vigour, old ideas which had for the last few decades been put to the side. Under novel labels such as ‘New Classical’ and ‘New Keynesian’ theory, explanations of unemployment being simply of a voluntary or merely frictional character were reasserted, attracted sympathetic listeners and soon found their way into the burgeoning crop of macro textbooks coming on the market. Over the years distinctive features of the Keynes theory – such as the concepts of involuntary unemployment, of the marginal efficiency of capital as distinct from the marginal productivity of capital, of uncertainty as something different from mathematically measurable risk, and the understanding that the macro economy contained within itself, even in the long run, no reliable self-righting mechanism to guarantee the automatic establishment of full employment – tended to slip out of the mainstream picture. Indeed, more than that: the Keynes theory is frequently misrepresented – it is typically asserted that an assumption of wage-stickiness is the critical factor differentiating the Keynes theory from the classical theory. Scholars who should have known better have been all too ready to adopt the old classical labour market theory of unemployment as embodied in the AS curve, apparently seeing the AD/AS model as a convenient and acceptable device for allowing analysis to be extended beyond the fix-price world of IS/LM. The upshot is that mainstream teaching of macroeconomic theory is today typically propounding a view of the working of the economy which is a very long way from the vision presented in the *General Theory* or from the conventional wisdom of the immediate post-war years, but strikingly similar to views current long ago, before the ‘Keynesian Revolution’. It is not going too far to say that the practical common-sense of the Keynesian perspective has (at least in some not un-influential quarters) been replaced by irrelevance and fantasy.¹⁸

It would not be a matter for concern if the eclipse of the Keynesian theory had occurred for the reason that it had been assessed and found wanting on empirical or theoretical grounds. *But that is not what has happened.* Rather it seems that we have drifted into this position through failure (when, with the revival of the old orthodoxy, it really mattered) to understand or remember just what differentiates the Keynesian from the classical theory. An inability to:

‘grasp the true nature of Keynes’s departure from orthodoxy is demonstrated by efforts to supplant the Keynes theory by conceptions such as “rational expectations”, the “New Classical Theory”, and the (so-called) “New Keynesian Theory”’: these in fact miss the Keynesian point.’

In the light of recent experience involving financial crisis, collapsing business and consumer confidence and persisting unemployment reaching levels reminiscent of the inter-war years, the Keynesian approach may, we suggest, be seen as somewhat more relevant than a theory which blames the unemployed for their plight and confidently asserts that the macroeconomic system possesses a strong natural propensity to establish a comfortable situation of full employment. As Keynes said long ago, such a doctrine is ‘misleading and disastrous’ if taken as a guide to macroeconomic policy in the real world.

¹⁸ ‘Fantasy’? – while the traditional classical theory held that workers could be unemployed because they were unwisely demanding wages too high to permit their employment, some present-day New Classical analyses present the thesis that when workers are not in employment they are actually in their preferred position: they are supposed deliberately to have chosen leisure over work.

References

- Abel, A B, Bernanke, B S and Croushore, D (2011), *Macroeconomics* (7th edition). Harlow, Essex: Pearson Education.
- Barro, R J and Grilli, V (1994), *European Macroeconomics*. Basingstoke: Macmillan.
- Blanchard, O (2011), *Macroeconomics* (5th – updated – edition) Boston, MA: Pearson Education.
- Burda, M and Wyplosz, C (2013), *Macroeconomics: A European Text* (6th edition). Oxford: Oxford University Press.
- Capaldo, J (2013), 'Fiscal Austerity and the Euro Crisis', *GDAE Policy Brief No 13-01*, Nov. 2013. Medford, MA: Tufts University.
- Colander, D C (1995), 'The Stories We Tell: A Reconsideration of AS/AD Analysis', *Journal of Economic Perspectives*, 9 (3), pp.169-188.
- Dornbusch, R, Fischer, S and Startz, R, (2011), *Macroeconomics* (11th edition). New York: McGraw-Hill.
- Friedman, M (1975), *Unemployment versus Inflation?* London: Institute of Economic Affairs, Occasional Paper 44.
- Froyen, R T (2013), *Macroeconomics: Theories and Policies* (10th edition). Upper Saddle River, NJ: Pearson Education International.
- Gordon, R J (2012), *Macroeconomics* (10th edition). London: Pearson Education.
- Grieve, R H (1998), 'Aggregate Demand and Supply: Two Into One Won't Go', in B B Rao (ed.) (1998), pp.83-94.
- Grieve, R H (2010), 'Time to Ditch AD-AS?' *Review of Radical Political Economics*, 42 (3), pp.315-320.
- Hansen, A H (1953), *A Guide to Keynes*. New York: McGraw-hill.
- Hicks, J R (1937), 'Mr Keynes and the 'Classics': A Suggested Interpretation', *Econometrica*, 5 (2), pp.147-159.
- Kennedy, P (1998), 'Defending ADAS: A Perspective on the ADAS Controversy', in B B Rao (ed.) (1998), pp.95-106.
- Keynes, J M (1936), *The General Theory of Employment, Interest and Money*. London: Macmillan.
- Keynes, J M (1939), 'Relative Movements of Real Wages and Output', *Economic Journal*, 49, March, pp.34-51.
- Krugman, P and Wells R (2013), *Macroeconomics* (3rd edition). New York: Worth Publishers.
- Makin, J M (2006), 'Japan Gingerly Exits Deflation', *Economic Outlook*, AEI Online.
- Mankiw, N G (2013) *Macroeconomics* (8th edition). New York: Worth Publishers.
- Moseley, F (2010), 'Criticisms of Aggregate Demand and Supply and Mankiw's Presentation', *Review of Radical Political Economics*, 42 (3), pp.308-314.
- Moseley, F and Grieve, R H (2010), 'Reply to Scarth', *Review of Radical Political Economics*, 42 (3), pp.327-329.
- Nevile, J W and Rao, B B (1996), 'The Use and Abuse of Aggregate Demand and Supply Functions', *The Manchester School*, 64 (2), pp.189-207.
- Patinkin, D (1959), 'Keynesian Economics Rehabilitated: A Rejoinder to Professor Hicks', *Economic Journal*, 69 (2), pp.582-587.
- Pigou, A C (1933), *The Theory of Unemployment*. London: Frank Cass.

Rao, B B (1991), 'What is the Matter with Aggregate Demand and Supply?' *Australian Economic Papers*, 30 (4), pp.264-277.

Rao, B B (ed.) 1998), *Aggregate Demand and Supply: A Critique of Orthodox Macroeconomic Modelling*. Basingstoke: Macmillan.

Rao, B B (2007), 'The Nature of the AD/AS Model Based on the IS/LM Model', *Cambridge Journal of Economics*, 31 (2), pp.413-422.

Scarth, W (2010), 'Aggregate Demand-Supply Analysis and Its Critics: An Evaluation of the Controversy', *Review of Radical Political Economics*, 42, (3), pp.321-32.

Author contact: roygrieve@btinternet.com

SUGGESTED CITATION:

Roy H. Grieve, "'Right back where we started from': from 'the Classics' to Keynes, and back again", *real-world economics review*, issue no. 68, 21 August 2014, pp. 41-61,
<http://www.paecon.net/PAERreview/issue68/Grieve68.pdf>

You may post and read comments on this paper at <http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

Demand theory is founded on errors

Jonathan Barzilai [Dalhousie University, Canada]

Copyright: Jonathan Barzilai, 2014

You may post comments on this paper at

<http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

1. The issue is applicability of mathematical operations

The applicability of the operations of algebra and calculus is a foundations-of-science problem. These operations have been applied incorrectly and where they are not applicable in microeconomic theory, the theory of games, decision theory and throughout the social sciences because the conditions for applicability of mathematical operations have not been identified in the literature. The applicability of these operations, in particular in demand theory, is founded on errors that are analyzed in detail in Barzilai [2 and 3].

In a recent paper [7], Katzner, whose work [6] contains these errors, says that:

“Jonathan Barzilai, in a paper entitled, ‘Inapplicable Operations on Ordinal, Cardinal, and Expected Utility’ has raised important issues regarding ordinal utility, and correctly clarified the meaning of the general notion of ordinality in terms of the mathematical theory of measurement. In that process, he has also subjected the traditional theory of consumer demand to serious attack.”

Having said that the meaning of the notion of ordinality has been correctly clarified, Katzner proceeds to obfuscate it by claiming a second notion of ordinal utility:

“Barzilai’s assault on traditional consumer theory, which is based on the mathematical theory of measurement, is useful because it brings to the fore the fact that, for economists, there is a second notion of ordinal utility, older than and independent of the mathematical-theory-of-measurement concept, and which is the relevant one for the traditional theory of consumer demand. That older approach seems to have had widespread acceptance among economists before the newer mathematical approach was known to them.”

The following should be noted:

1. My analysis is not based on the mathematical theory of measurement. As Katzner notes, the title of my paper is “Inapplicable Operations on Ordinal, Cardinal, and Expected Utility.” The subject of applicability of operations does not appear in measurement theory (see [7, 10, and 11]). Furthermore, in Section 3.8 of [2] I show that the mathematical theory of measurement is flawed and is of no scientific value.
2. Consumer preference is preference whether it is studied by economists, psychologists, mathematicians, or physicists. Preference under any name, including utility, value, “wants and desires,” tastes, or ophelimity is preference (see Section 3.9.4 of [2] for a detailed discussion of this issue). The notion of a different kind of preference for demand theory has no logical basis.

3. The mathematical theory of utility which is relevant to the traditional theory of consumer demand is subject to the same mathematical principles that apply to any other mathematical theory. The fact that there is widespread acceptance among economists of older incorrect notions of the mathematical theory of utility indicates an urgent need to correct these notions. Errors whose acceptance is widespread need to be corrected rather than defended.
4. The marginal utilities which are relevant to the traditional theory of consumer demand are partial derivatives of utility functions. Derivatives are concepts of differential calculus. There is no second notion of differential calculus which is the relevant one for the traditional theory of consumer demand. Elementary calculus errors by Hicks, Samuelson, and their followers are analyzed in detail in [2 and 3]. Katzner is defending the widespread misapplication of differential calculus in the traditional theory of consumer demand.

2. More on Hicks's and Samuelson's errors

The operations of calculus, including differentiation, are carried out in a vector space (see e.g. Dieudonne [4]). Vector spaces and the operations of calculus are quantitative concepts (for formal definitions see [2, §3.7]). It follows that the notion of "non-quantitative calculus" is a contradiction in terms yet, according to Hicks [5, p. 19], the operation of differentiation is applicable on utility functions that he has "purged" from quantitative concepts. Hicks and all economists who reject "all concepts which are tainted by quantitative utility" are rejecting the application of calculus in utility theory and thereby the very concept of marginal utility. Calculus is quantitative, differentiation is quantitative, and marginal utility, which is a derivative, *is* a quantitative concept. The notion of non-quantitative differentiation is unique to microeconomics.

The Hicksian purge applies to *all* concepts of quantitative utility of *any* kind. Furthermore, there is no support for Katzner's second kind of ordinal utility in the literature. Repeating Hicks's ordinal utility error, Samuelson correctly gives *the only possible definition of the only kind of ordinal utility* in Equations (6-8) of [12, p. 94] which he verbally describes [12, p. 91] as "ordinal preference, involving "more" or "less" but not "how much," but then he incorrectly claims that only ordinal preference is required for the analysis of consumer's behavior. Also note Samuelson's use of *preference* as synonymous to *utility*.

In addition, Hicks [5], Samuelson [12], Mas-Colell et al. [9], and all authors who claim that ordinal information is sufficient for the existence of utility derivatives, including Katzner [6], rely on an incorrect application of the Implicit Function Theorem of calculus. The onus is on these authors to establish that the assumptions of this theorem (see e.g. Apostol [1, p. 147]) are satisfied. Applying the Implicit Function Theorem where the conditions for its applicability are not satisfied is an elementary error in consumer demand theory.

Samuelson correctly says (see [12, p. 94, Equation (9)]) that any monotone increasing transformation of an ordinal utility function is an equivalent ordinal utility function, but the rest of his argument fails on the infinitely many non-differentiable monotone increasing functions that cannot be differentiated. Differentiating these non-differentiable transformations is an elementary error. Moreover, Samuelson's faulty argument applies verbatim to the case where

no ordinal information on the indifference surfaces is available (the numerical value of the utility of x equals the numerical value of the utility of y if and only if the consumer is indifferent between x and y). This implies the absurd claim that the quantitative tools of differential calculus apply on utility functions where the only available information is whether $u(x)$ does or does not equal $u(y)$. The ordinal utility claim, which is based on the same errors, is just as absurd. The notion of differentiating ordinal functions has no counterpart in science — vector space operations are not applicable on ordinal data and ordinal functions are not differentiable. Physics (and mathematics) should be rewritten if ordinal information is sufficient for the application of differential calculus.

Finally, if the partial derivative of a utility function with respect to one of its variables does not exist, the assumptions of the Implicit Function Theorem are not satisfied and this theorem cannot be employed to “prove” that although the derivatives do not exist, their ratios do exist (cf. Hicks [5, p.19]). This, too, is an error.

3. Summary

The claim that there is a second notion of ordinal utility on which a second kind of mathematics applies is untenable. Katzner cannot ignore the fundamental issue of the conditions for applicability of the operations of algebra and calculus. These operations cannot be applied where the conditions for applying them are not satisfied. He cannot ignore the counter-examples and the detailed analysis of the errors committed by Hicks and Samuelson in Section 3.4 of [2]. And he must show that the conditions for applying the Implicit Function Theorem are satisfied where they are used in demand theory. Demand theory’s errors should be corrected, not defended.

References

- [1] Tom M. Apostol, *Mathematical Analysis: Modern Approach to Advanced Calculus*, Addison-Wesley, 1957.
- [2] Jonathan Barzilai, “Preference Function Modeling: The Mathematical Foundations of Decision Theory,” in *Trends in Multiple Criteria Decision Analysis*, Matthias Ehrgott, José Rui Figueira, Salvatore Greco (Eds.), Springer, pp. 57—86, 2010. Preprint at www.ScientificMetrics.com
- [3] Jonathan Barzilai, “Inapplicable Operations on Ordinal, Cardinal, and Expected Utility,” *Real-World Economic Review*, No. 63, pp. 98—103, 25 March 2013.
- [4] Jean A. Dieudonne, *Foundations of Modern Analysis*, Academic Press, 1960.
- [5] John R. Hicks, *Value and Capital*, Second Edition, Oxford University Press, 1946.
- [6] Donald W. Katzner, *An Introduction to the Economic Theory of Market Behavior: Microeconomics from a Walrasian Perspective*, E. Elgar, 2006.
- [7] Donald W. Katzner, “Ordinal Utility and the Traditional Theory of Consumer Demand,” *Real-world Economics Review*, No. 67, pp. 130-136, May 2014.
- [8] David H. Krantz, R. Duncan Luce, Patrick Suppes and Amos Tversky, *Foundations of Measurement*, Vol. 1, Academic Press, 1971.
- [9] Andreu Mas-Colell, Michael D. Whinston and Jerry R. Green, *Microeconomic Theory*, Oxford University Press, 1995.

- [10] Louis Narens, *Abstract Measurement Theory*, MIT Press, 1985.
- [11] Johann Pfanzagl, *Theory of Measurement*, Wiley, 1968.
- [12] Paul A. Samuelson, *Foundations of Economic Analysis*, Harvard University Press, 1947.

Author contact: barzilai@dal.ca

SUGGESTED CITATION:

Jonathan Barzilai, "Demand theory is founded on errors", *real-world economics review*, issue no. 68, 21 August 2014, pp.62-65, <http://www.paecon.net/PAEReview/issue68/Barzilai68.pdf>

You may post and read comments on this paper at <http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

The Central Bank with an expanded role in a purely electronic monetary system

Trond Andresen [The Norwegian University of Science and Technology, Norway]

Copyright: Trond Andresen, 2014

You may post comments on this paper at

<http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

Abstract

Physical currency (bills and coins) is being phased out as an important means of exchange both in developed and developing countries. Transactions are increasingly done by debit card, computer, and mobile phone. This technologically driven process opens up some very useful possibilities, among these new and – for society – beneficial roles for the Central Bank. The paper assumes a scenario where the country in question issues its own currency, and all money is “electronic” – no bills and coins. This gives an extra impetus to the *sovereign money* solution; all deposits are at the Central Bank.

The paper also argues that in such a system – where banks are not allowed to create “credit money” when issuing loans (in this resembling the “100% reserve” solution supported by many reformers) – the economy need not, in spite of this, be “starved” of credit for investment – a warning that is not only sounded by the defenders of today’s financial system, but also by many of its critics. This goal might be achieved by the unconventional trick of letting *commercial banks create the needed sovereign money at the Central Bank* for their lending.

A third point of the paper is to argue that simplification of the financial system should be a goal in itself.

JEL codes B50, E5, E40, E42, E44, E58, G20, G28, H12, H62

1. Introduction

The technological development process that allows electronic transaction instead of exchanges using physical currency, has the same merciless and irreversible character as the advent of the electronic calculator in the 70s and digital photography in the 90s: it meant the unavoidable death of the slide rule (then) and photographic film (more recently). Based on the nature of technological innovations and the market economy’s exploitation of such, we may predict the death of physical currency; bills and coins. It is probably a question of when, not if, this will take place. This paper will discuss some positive possibilities for reform of the financial and monetary system that emerge as a side effect of the unstoppable advances of technology in this field.

A modern financial system consists of a Central Bank (CB) and an extensive network of private financial units. The role of a CB has up to this day been as an interest-rate setter behind the scenes and – in crisis – “lender of last resort” for the network of private licensed (“commercial”) banks and non-bank financial institutions (NBFIs).

The commercial bank network has historically been quite dense, with branches of competing banks within a reasonable distance from customers. The reasons for this geographical diversity has been twofold:

1. Handling deposit accounts and receiving or furnishing customers with physical currency.
2. Vetting potential borrowers and extending loans.

With the advent of electronic transactions (PC, debit card and, lately, mobile phone) the need for a dense network of branches has decreased, and commercial banks have started the process of closing down an increasing share of these. If we envisage an expected future without physical currency, the first point above will disappear as a reason for having bank branches. What remains is the second point, the need for offices to handle decisions about loan applications, which to a fair degree will be best handled by personnel having local and/or specialised knowledge. Except for this, most decisions can be taken at a bank's central office.

So, simply because of no more need for branches to acquire or deposit physical money – for purely technological reasons, not society's economic policy considerations – it becomes feasible for all "agents" (persons, firms, banks) to only have their checking accounts directly at the CB. Then one may dispense with bank credit money and let all money in circulation be *base money (high-powered money; HPM)*. For the public this means that their deposits are completely safe, and in that sense it matches the 1930ies "Chicago plan" and its "100% money" proposal. But it goes further, because in the 100% money plan, banks would hold people's deposits (although fully covered by the banks' deposits at the CB) and furnish them with physical currency, while in the above electronic money scenario liquid deposits only exist directly at the CB. This is the *sovereign money* alternative (Huber 2014). Technological possibilities today give a new impetus to this alternative, which has until now been promoted based on political economy arguments only.

We will from now on describe and argue for a banking scenario based on the sovereign money alternative, but with only electronic currency, using the acronym "ESMA – Electronic Sovereign Money Alternative".

We will first address an argument raised against both 100% money and sovereign money: "*when banks are not allowed to create credit money, the economy will suffer because of lack of credit*".

2. The heterodoxy is partly hostile to 100% money or sovereign money

The famous pre-WWII Fisher et al *Chicago Plan*, was more recently re-examined (Benes & Kumhof , 2012), and in conclusion supported, where they write in the abstract:

At the height of the Great Depression a number of leading U.S. economists advanced a proposal for monetary reform that became known as the Chicago Plan. It envisaged the separation of the monetary and credit functions of the banking system, by requiring 100% reserve backing for deposits. Irving Fisher (1936) claimed the following advantages for this plan: (1) Much better control of a major source of business cycle fluctuations, sudden increases and contractions of bank credit and of the supply of bank-created money. (2) Complete elimination of bank runs. (3) Dramatic reduction of the (net) public debt. (4) Dramatic reduction of private debt, as money creation no longer requires simultaneous debt creation. We study these claims by embedding a comprehensive and carefully calibrated model of the banking system in a DSGE model of the U.S. economy. We find support for all four of Fisher's claims. Furthermore, output gains approach 10 percent, and steady state inflation can drop to zero without posing problems for the conduct of monetary policy.

Ann Pettifor disagrees (Pettifor, 2013), and argues (page 20) that 100% reserve banking will lead to lack of credit:

The Kumhof and Benes proposal is indeed based on the monetarist ideas of the Chicago School, one that seeks to limit the quantity of money, and that would restore the role of banks to intermediaries between savers and borrowers. Only now the proposal is to eclipse the role of the private sector altogether, and only allow lending backed by a 100% reserve requirement. In other words, all banks or lenders would first have to mobilise 100% of the funds needed for lending. This would massively constrain the availability of credit. (...)

Limiting the quantity of credit is certainly one way of limiting employment. Thus monetarist theory and policies both tolerated and sustained a massive rise in unemployment in the 1930s and 1980s. The Kumhof and Benes proposal is no more than a revival of these policies: the 'barbaric relic' that was the gold standard.

Pettifor is hostile to the 100% reserve concept – and her platform is anti-neoliberal¹. She is not alone in this; many central authors in the heterodox Modern Monetary Theory (MMT) and/or Post Keynesian camps share her position. One of these is Jan Kregel, who describes and supports Hyman Minsky's critique of what he termed "narrow banking" (this corresponds to banks subjected to a 100% reserve requirement) in a paper that argues along similar lines (Kregel, 2012):

In the absence of a large government sector to support incomes, liabilities used to finance investment could not be validated in a narrow bank holding company structure. But, even more important, it would be impossible in such a system for banks to act as the handmaidens to innovation and creative destruction by providing entrepreneurs the purchasing power necessary for them to appropriate the assets required for their innovative investments.

Emphasising the need for easy access to credit for "Schumpeterian creative destruction", Kregel argues that if banks are not allowed to create extra money when lending, what is left:

... is not a bank, but simply a safe house or piggy bank for government issues of coin and currency.

Kregel, however, points to a possible solution to – or amelioration of – lack of capital for investment:

In the absence of private sector "liquidity" creation, the central bank would have to provide financing for private sector investment trust liabilities, or a government development bank could finance innovation through the issue of debt monetized by the central bank. (...) such a system would have to combine Keynes's idea of the "socialisation of investment" with the "socialisation" of the transactions-and-payments system. (...)

¹ I share this general anti-neoliberal position. And I will not defend DSGE – the main tool for neoclassical modeling – used in the Kumhof and Benes proposal, and all other details there. But I agree with them about the benefits of (some sort of) a "100% reserve" system.

the real problem that must be solved lies in the way that regulation governs the provision of liquidity in the financial system.

This paper proposes regulation that may achieve what Kregel suggests, but by the unconventional road of letting commercial (licensed, “narrow”) banks *decide on creation of HPM at the Central Bank*. This will be explained in the following.

3. Ample credit lines for banks at the Central Bank

Consider an economy where all money is base money (HPM): Could banks not – if they mean they have a worthwhile and fairly safe lending opportunity – just borrow HPM from the Central Bank and re-lend it at a somewhat higher interest rate? This is in contrast to today’s state of affairs where credit money is created directly through bank lending, completely dominating money growth.

Such bank borrowing from the CB implies that HPM will grow as an effect of this, not only through government deficit spending (if we follow the MMT advice of financing government deficits by directly “borrowing” from its CB, instead of selling bonds to banks and the public). So in such a scenario, some money will not only be *spent* into circulation (government deficit), some will also be *lent* into circulation (via banks). *But all of it will be HPM*: The amount of extra money created and subsequently put into circulation due to bank borrowing from the CB will be completely safe, not credit money which carries some risk.

Banks could also gather money for their lending by selling bonds or offer time deposits to the public. Then the amount of extra money created by bank borrowing from the CB would constitute only a share of new loans given. But we will argue that this alternative, which puts the saver at some risk, is not necessary.

To sum up at this stage, our ESMA scenario assumes that all government deficits are financed by “loans” from the CB (“loans” in quotation marks, if we regard the CB a tool of the government – following MMT), and that all bank lending is financed by corresponding loans to banks from the CB. Banks will resemble “franchisees” of the CB, living off the difference on interests in and out. The role left for banks is to be pure intermediaries.

3.1 A quite harmless “bailout”

With ESMA, if a bank defaults, the bank owners would lose all their equity. Hence, any bank would have a much stronger incentive for responsible behaviour than in today’s environment. And there would be no need or reason for society to step in with bailouts. This will be a credible threat that banks’ owners have to take seriously.

The impact on the CB would be much less dramatic. The only “loss” to the CB and society when a bank defaulted on loans from the CB, would be that the corresponding HPM that was supposed to be destroyed though repayments with interest to the CB, *remained in circulation*. Instead of society increasing taxes to pay for bailing out a bank, money that was already in circulation would not be retired. The effect of this would then be spread thinly over society as a whole. At worst this would give a small impulse for inflation. We may contrast this small disadvantage with the big advantage of directly hitting the bank owners in this proposal; they

would lose their assets. And this no-bailout system should – even with interbank lending – be 100% robust in a systemic sense.

Based on the above, licensed banks can be given very ample credit lines to the CB, at reasonably low rates. This is the main reason that society's need for credit should not be constrained in a damaging way.

That said, banks should not be allowed unlimited borrowing from the CB. This is discussed in the next subsection.

3.2 A BIS-type capital adequacy constraint is feasible

In today's regulatory environment, banks are to a decreasing degree reserve constrained in different countries. The trend is towards implementing Bank for International Settlements-type regulation that only sets a lower threshold for commercial banks' capital adequacy. We will discuss this based on a simplified representation of this regulatory framework that is taken from (Andresen, 2010), and in the next stage we suggest a very similar capital adequacy requirement tailored for the ESMA scenario.

We define:

$M(t)$ = deposits = money stock [\$], the bank's liability. We abstract from notes and coins, and from interbank lending which has no important bearing on the points to be made.

$D(t)$ = loans from bank [\$].

$R(t)$ = reserves = a bank's deposit with the CB = high-powered money (HPM) [\$]. We assume that $R > 0$. The Bank's total financial assets are now $D + R$, where D = loans as before.

k_0 = the required minimum capital/asset ratio []. This entity is dimensionless, thus the [].

k = the actual capital/asset ratio [].

Variables' dependence on time (t) is from now on not shown. Remembering the Basel rule that risk weights shall only apply in the denominator and that reserves R carry zero risk weight, we get the requirement

$$\frac{D+R-M}{D+0\cdot R} = \frac{D+R-M}{D} = k \geq k_0 \quad (1)$$

What happens when the bank extends a loan ΔD ?

Since both D (the bank's asset side) and M (liabilities side) increase with ΔD , the nominator remains the same, while the denominator increases with ΔD . The result is a fall in k towards k_0 . This may imply a restriction on further lending, and that is the purpose of the regulation: to achieve some minimum robustness against insolvency.

Now to the ESMA scenario. We additionally define:

$D_{CB}(t)$ = the bank's debt to the CB [\$].

We now suggest the requirement:

$$\frac{D+R-D_{CB}}{D+0\cdot R} = k \geq k_0 \quad (2)$$

The liability towards the CB here plays the same role as today's bank liability M towards its depositors. When the bank extends a loan ΔD , the situation for the bank afterwards is:

$$\frac{D+\Delta D+R-(D_{CB}+\Delta D)}{D+\Delta D} = \frac{D+R-D_{CB}}{D+\Delta D} = k \quad (3)$$

The ratio k is decreased in the same way as today. We thus achieve a regulatory constraint on banks' lending behaviour which formally is quite similar to that given by the current regulations, except for one difference: in today's environment we have a two-way relationship (the bank and the borrower), while the relationship in the ESMA scenario becomes *triangular*: the bank, the CB, and the borrower. The bank gets a claim ΔD on its borrowing customer, the CB gets the same extra claim on the bank. The balance sheet of the bank customer increases with ΔD on the customer's liability side at the bank, while the customer's checking account at the CB is credited with ΔD . But the change from a dual to this triangular relationship does not impact on the efficacy of the suggested ESMA capital adequacy rule.

4. The CB as a lending and savings hub for society

So far about CB lending to banks. On the *savings* side, the CB can – due to the information technology revolution discussed above – offer individual accounts not only for banks, but for *all* agents: citizens and firms; both a checking account and a spectrum of time deposits yielding different rates, payment profiles and durations. Since individual depositors' money at the CB – whether from persons or businesses – would be completely risk-free, a checking account there should yield zero interest. But such accounts could be cost-free for the user, considered part of a modern welfare state's shared free infrastructure, like healthcare and schools.

By tweaking interest rates on its lending, the CB can ensure that banks get the necessary incentive to lend, by a sufficient difference between bank lending rates and (for them) CB borrowing rates. On the other side, by offering sufficient rates on its spectrum of time deposits, the CB can withdraw money from aggregate demand, from banks, firms and individuals.

4.1 A simplified and understandable system

For a modern monetary economy to function well, there needs to be democratic discussion among the public of the workings of the financial and monetary system. Such discussion is simply not feasible today due to the enormous complexity of finance. One reason for this

complexity is explained as “innovation” to cater to additional “needs of the market” as the defenders of today’s state of affairs say. Another reason which is much less discussed is that this complexity makes finance impenetrable to the public and lawmakers, so that they are cowed and leave the financial sector alone and give up on reforms. Enter the ESMA, which has the added benefit of making the financial and monetary system much simpler, and thus accessible for public discussion, control and reform.

We have until now discussed the role of commercial (licensed) banks with the ESMA. But in what way – maybe negatively – will such simplification impact some of the large NBFIs that also constitute the modern financial sector? Let us take pension funds as an example. Abstracting from assets overseas, with ESMA they will have a portfolio consisting of – like today – company stocks and property (which should and can be regulated much better to avoid bubbles, but this is not a topic for this paper), *and large assets in the form of time deposits or bonds*² at the CB. These assets are again completely safe, which is of course a benefit to the pension savers. If a fund wishes for higher returns and accepts the ensuing higher risks, it can compose a portfolio with a smaller CB savings component. There is no important loss of flexibility for a pension fund with the ESMA.

5. Concluding remarks

In the proposed ESMA, the government runs more or less persistent fiscal deficits – the normal state of affairs recommended by MMT – by “borrowing” from the CB. There is no need to issue government bonds to cover the deficit. The CB only offers a spectrum of shorter term securities as an adjustable tool for interest rate control.

The interest rate is essentially controlled by the CB, as it is now. Money supply and debt growth may be better controlled, to follow physical growth harmoniously. Banks cannot net create money by themselves, like they do today³. But most important is that their debt creation will be less irresponsible, because of the effective threat of loss to bank owners.

There will be no lack of flexibility and agility in this scenario however, since licenced banks will enjoy ample credit lines to the Central Bank to access the necessary HPM in time, when giving a new loan. They can also borrow from other banks (but allowing interbank lending is not a prerequisite for the ESMA). By this, they can grant a loan just as easily and quickly as they do today. The amount of net HPM created through government deficit spending plus that created through banks’ borrowing from the Central Bank, can be controlled by fiscal and monetary policy to be sufficient, among other things so that there is no harmful lack of credit.

And banks that currently have too much HPM and see too few lending opportunities can save at the Central Bank, using the spectrum of paper offered by the Central Bank for that purpose.

² One issue which will be left out here, is whether the CB should offer only non-transferable time deposits, or also sell a spectrum of securities, which – as opposed to time deposits – can be traded. The choice here does not impact much on the workings of the ESMA.

³ The mathematical mechanics of this is discussed in (Andresen, 2010). An essential result there is that the growth rate of debt and credit money is inversely proportional to the minimum required capital adequacy ratio, called k_0 above. This gives fairly steep growth when k_0 is less than 10%, and banks for profitability reasons wish to stay at this limit.

References

Andresen, T. (2010). Basel rules, endogenous money growth, financial accumulation and debt crisis, *The Aftermath of The Crisis*, Centre of Full Employment and Equity, University of Newcastle, Newcastle, Australia, pp. 8–20.

URL http://www.itk.ntnu.no/ansatte/Andresen_Trond/econ/basel.pdf, accessed 28 May 2014.

Benes, J. & Kumhof, M. (2012). The Chicago Plan Revisited, *working paper*.

URL <https://www.imf.org/external/pubs/ft/wp/2012/wp12202.pdf>, accessed 28 May 2014.

Huber, J. (2014). Modern money theory and new currency theory - a comparative discussion, including an assessment of their relevance to monetary reform, *Real-world economics review* (66).

URL <http://www.paecon.net/PAERReview/issue66/Huber66.pdf>, accessed 28 May 2014.

Kregel, J. (2012). Minsky and the narrow banking proposal: No solution for financial reform, *Public Policy Brief*, *Levy Institute*.

URL <http://www.levyinstitute.org/publications/minsky-and-the-narrow-banking-proposal>, accessed 28 May 2014.

Pettifor, A. (2013). The power to create money 'out of thin air' - understanding capitalism's elastic production of money and moving on beyond Adam Smith and 'fractional reserve banking' - a review essay of Geoffrey Ingham's *Capitalism*, *Prime economics*.

URL <http://www.primeeconomics.org/wp-content/uploads/2013/01/The-power-to-create-money-out-of-thin-air5.pdf>, accessed 28 May 2014.

Author contact: trond.andresen@itk.ntnu.no

SUGGESTED CITATION:

Trond Andresen, "The Central Bank with an expanded role in a purely electronic monetary system", *real-world economics review*, issue no. 68, 21 August 2014, pp. 66-73,

<http://www.paecon.net/PAERReview/issue68/Andresen68.pdf>

You may post and read comments on this paper at <http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

Financialization, income distribution, and social justice: recent German and American experience

Robert R. Locke¹ [Germany/USA]

Copyright: Robert R. Locke, 2014

You may post comments on this paper at

<http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

Although many economists and business writers have discussed increased financialization since the 1970s, they have paid scant attention to the impact this change may have had on the distribution of incomes in Western economies. This paper compares and contrasts in this respect American financialization with German over the past half-century. According to Petra Dünhaupt, financialization in the two countries differed: “In the US, the important shift towards financialization occurred in the early 1980s, ...in Germany the process of financialization started much later – in the beginning of the 1990s – and followed a much more gradual transition.” (Dünhaupt, 2012, 1) The analysis is pursued historically, on the grounds that an understanding of the financialization of the German economy requires an investigation of intergenerational institutional legacies.

The financialization referred to here can be described as the transition from management capitalism to finance capitalism. More specifically, it is the change from viewing a business as a vehicle for earning “returns on investment . . . based on the value created by productive enterprise” to viewing a business “as assets to be bought and sold for maximizing profits through financial strategies.” (Ball & Appelbaum, 2) Large-scale industrial organizations in which the internal “visible hand” of management orchestrated productive activities formerly coordinated through the external “invisible hand” of market transactions first emerged in the United States in the 19th century. By the 1920s, many of these organizations, as famously described by Alfred D. Chandler, Jr. in his path-breaking book *The Visible Hand*, 1977, (Locke & Spender, 2011) had evolved into complex multi-divisional enterprises that arguably represent the pinnacle of management capitalism’s development. Indeed, large managed enterprises have flourished in the global economy from the early 20th century to today. But changes in financial markets, financial institutions, and management compensation after the 1960s increasingly shifted the attention of managers from producing and selling products and non-financial services to seeking returns from financial activities. These changes from management capitalism to finance capitalism had profound consequences for the distribution of incomes. However, the fairness of that distribution varied greatly from nation to nation, depending on their institutional and governance forms and how democratically members of each society could alter those forms. Here we particularly examine how institutional and governance differences shaped outcomes in Germany and the United States.

Financialization and executive compensation

Epstein defines “financialization” as the “increasing importance of financial markets, financial motives, and financial institutions, and financial elites in the operation of the economy and its

¹ I would like to thank H. Thomas Johnson for his critique of my original comments about financialization.

governing institutions, both at the national and international level” (Epstein 2005, 1). Dünhaupt posits five ways that financialization has affected the compensation of executives:

1. It shifted the basis of enterprise finance from banks to capital markets;
2. It reinvigorated the “rentier” class that had been on the decline by creating institutional investors (e.g., pension funds) that base investment decisions solely on stock prices and short-term return on investment;
3. It linked financial trading to new financial institutions (e.g., investment banks, hedge funds, and private equity firms) and new financial instruments (e.g., derivatives, stock options, and credit swaps);
4. It stressed profit-making through financial activities instead of through real productive activity;
5. Under the guise of increasing share-holder value in a firm, it subordinated the interests of stockholders as governing agents of nonfinancial firms to those of directors (and, implicitly, those of Wall Street analysts, investment bankers, and large investors). In effect, this made directors, not stockholders, the chief beneficiaries of financialization (Dünhaupt, 2011, 10; Locke, 2012; Ball & Appelbaum, 3).

The fifth point raises an interesting issue. The financialization of firm governance is often equated with stockholder primacy because decision-making still resides in the CEO and the board of directors. Thus, because the stockholders choose the CEO and board it suggests that they remain in charge. This, however, is not exactly what happened in large US firms. Through control of the proxy process, incumbent CEOs have come to nominate their own candidates for board memberships, thereby making the boards director-selected instead of shareholder-selected entities. According to Stephen Bainbridge “director primacy” has placed “power and the right to exercise decision- making neither in shareholders nor the managers, but in the Board of Directors,” who have claimed the right to assure the “efficient maximization of shareholders’ residual claims,” without the shareholders in effect being able to control the board (Bainbridge, 2006, 1).

If CEOs have escaped stockholder control through their *de facto* power to appoint directors, the rise of the institutional fund managers gives even further power to the CEOs. The institutions that employ these managers, primarily public and private pension funds, by 2000 came to own almost fifty percent of the equity of American corporations. Approximately fifty percent of Americans either owned stock individually or, more typically, had an ownership or retirement interest in these fiduciary institutions. The institutional fund managers did not threaten director primacy or CEO control because they could step in and micromanage firms in which they held stock. Institutional fund managers have in fact honed few of the skill-sets necessary to replace errant corporate managers. These fund managers live in the investment world, operate by its rules, and have little knowledge about how actually to manage firms in which they place investments. Laurence Mitchell stressed this point in *The Speculation Economy* when he described how the institutional investment managers slavishly follow the Capital Asset Pricing Model (CAPM) in their investment calculations:

“The product of a regression analysis called *beta*, CAPM allows investors to build the kinds of potentially lower-risk, higher-return portfolios ...described by [the Nobel Prize winning economist] Markowitz, based solely upon a narrow range of information about the stock. The business itself matters little, if at all. All an investor needs is *beta*. No balance sheet, no profit and loss statement, no cash flow information, no management analysis of its performance and

plans, no sense of corporate direction, no knowledge of what is on its research and development pipeline, no need even to know what products the corporation makes or what services it provides. Just *beta*. The stock is virtually independent of the corporation that issued it. CAPM has been adopted and is daily used by countless stock analysts and institutional money managers. Almost every American who invests in the market through mutual funds or other institutional media has invested on the basis of CAPM” (Mitchell, 2008, 275).

Nonetheless, the rise of institutional investors has affected firm governance in two important ways. First, everybody knows that the incomes of the top one percent of Americans, in which category American CEOs belong, have increased dramatically in recent decades. If their percentage of wages is subtracted from labor’s share of seventy-one percent in the late 1970s (the bottom ninety-nine percent of wages earners), the percentage of income of the bottom ninety-nine percent declines as of 2005 by ten percent, which means that only the incomes of the top one percent grew, and did so substantially, in those decades. (Dünhaupt, 10) That the incomes of the richest have benefitted handsomely is common knowledge, but the fact that the growing gap between the top one percent and the bottom ninety-nine percent can be attributed almost exclusively to the financialization of CEO salaries through stock options is perhaps not so well known. Dünhaupt claims as much -- that the introduction of stock options into American CEO pay is solely responsible for increasing their share of total incomes from two percent in 2000 to eight percent in 2007 (2011, 19). She concludes that given the proximity of CEOs’ position to capital owners rather than to workers, the stock option is closer to capital income than to wage income and should be classified with the former, i.e., with financialization, rather than with earned wages. Her point has been reinforced during this era of financialization, since the view of labor that prevailed under managerial capitalism as a quasi-fixed asset or human capital changes under financialization to one of “labor being considered a variable cost to be minimized” (Ball & Appelbaum, 6).

Financialization of director salaries encourages those at the top, in their own interest, to adopt a short-term Wall Street focus when running their companies. Mitchell reports that almost eighty percent of more than four hundred chief financial officers in major American corporations recently surveyed, would have at least moderately mutilated their businesses in order to meet analysts’ quarterly profit estimates. Cutting the budgets for research and development, advertising and maintenance and delaying hiring and new projects are some of the long-term harms they would readily inflict on their corporations to achieve good short-term numbers. The same influence of financialization holds when CEOs cut costs by downward pressure on employee wages and the elimination of a firm’s legacy costs (pensions and benefits), policies that have been relentlessly pursued during and since the last two decades of the 20th century (Locke & Spender, 2011, 153-56). These cost-cutting measures fit the financialization view that labor is not a firm asset (human capital) but a variable cost to be minimized. Cutting labor costs has a favorable impact on the market, driving up a firm’s stock price and with it the firm directors’ incomes.

The second way financialization affects director primacy governance is much more constraining for top executives. If not haunted with fears that institutional investors will take over daily management of their firms, firm directors do fear that institutional investors might divest their holdings of a firm’s stock if the firm’s profits and stock valuations fall. As Mitchell put it:

“Failure to meet quarterly numbers almost always guarantees a punishing hit to the corporation’s stock price. The stock price drop might cut executive compensation based on stock options, attract lawsuits, bring out angry institutional investors ... and threaten executive job security, if it happened often enough. Indeed, the 2006 turnover rate of 118 percent on the New York Stock Exchange alone justifies their fears” (Mitchell, 2008, 1).

A final point, financialization also invaded US capitalism in more profound and institutional ways – through the spawning of venture capitalist firms, angel investor networks and IPOs, through the promotion of private equity buyouts, amalgamations, and other schemes of privatization that whet the appetites of the investor class and fill the wallets of their agents with lucrative commissions for dealmakers in hedge funds, private equity firms, and investment banks.

The intensity of the reaction to Thomas Piketty’s claim in his blockbuster book *Capital in the Twenty-First Century* (“If the rate of return on wealth \otimes is greater than the rate of growth, then wealth is likely to be ever more concentrated”) is explained by financialization. Most of the debate seems to be about what Piketty means by Capital i.e., that he was emphasizing financial not physical capital. Merijn Knibbe in his 1 May 2014 post on the Real-World Economic Blog” explains “When we want to analyze economic growth, we might want to use an index of the amount of physical capital. When we want to study inequality, it might be wiser to look at the liability side of the balance sheet (finance capital).” People study the financialization process because it has contributed so much to inequality. It is handled as an example of convergence theory: what began in America and the UK spread, especially after the fall of communism in 1990 to the rest of the world using the financial theories and managerial tools of US-UK finance capitalism. But there is evidence that places outside their orbit, with divergent business and banking traditions reacted, because of them, more successfully resisted financialization and the crisis it brought. One of the places is Germany (see David Ruccio’s 5 May 2014 post in *rwer*, “12-country 1975–2007 chart of share of income growth going to the 1%” for distribution differences).

Germans adopted elements of American financialization later (stock options were not allowed in Germany until 1998), and German economic institutions exuded a methodological communitarian ideal in contrast to US methodological individualism and materialism that fostered financialization processes (stock option remuneration, IPOs, private equity firm takeovers, director primacy governance, etc.) that drove a wider gap between top and bottom US incomes than German. The German institutions discussed are in the banking and educational systems.

Banking systems

When the Berlin Wall came down in 1989, European banking everywhere rested on three pillars: private commercial banks, public saving banks, and co-operative banks. Traditionally the big German private banks had operated in the “kingly merchant tradition,” where a firm retained a *Hausbank* and relations with it rested on trust, i.e., customers were not customers in the American sense but clients (Batiz-Lazo, Locke, & Müller, 2008). The new information technologies churning out of America allowed the flow of monies to increase dramatically and permitted investors everywhere to trade in rapidly created equity markets twenty-four hours a

day. Taking advantage of this technology and the expanding geographical opportunities accompanying the collapse of communism, American and UK financial houses rooted in equity markets began to promote the financialization of enterprise in Europe by facilitating mergers and acquisitions, debt management, and capital acquisition.

This British and American *Drang nach Osten* [push to the East] affected the investment business of major German private commercial banks in their own country; by 2004 they only transacted 38.3% of the German merger and acquisition business, 21.8% of the German equity market business, and 16.3% of the debt market business (*The Economist* 1.11. 2004, 82). J.P Morgan, Morgan Stanley and Goldman Sachs beat the German banks in their own backyard because it was an American kind of capitalism. According to *The Economist* (27.03.2004, 75) the position of German banks became so bad that a German agency, the Kreditanstalt für Wiederaufbau, thought it best in order to optimize results in the privatization of Deutsche Telekom to auction off large blocks of the company's shares through foreign investment banks, rather than through the investment bank arms of Deutsche Bank, Dresdner Bank and other German banks.

German private commercial banks decided that survival depended on the adoption of the new model. They moved onto the turf of American and British capitalism, began trading in securities and engaging in business consultancy. They also, following the UK and US banks, marketed new products and services. These included selling loan packages, credit cards, insurance, and organizing electronic banking through automated machines, and on-line services. Banks acted less as *Hausbanken* for large companies and held less of their clients stock in their portfolios (Lütz, 2000). They shifted from the kingly merchant tradition environment of trust in retail banking to one of persuasion, to letting impersonal market mechanisms set price and determine transactions.

In 1990, the business model of the second pillar, the European public savings banks, had five features. First they were "public," which meant they were "in a certain sense owned or sponsored and governed by some regional or local public body such as a city or a county or region." Second, they were organized under a public law regime. Third, they had "a dual objective: They were expected to support the local economy and the local people, and at the same time to operate according to common business rules and thus to be financially sustainable enterprises." Fourth, they had to adhere to the "so-called regional principle, which restricts the operations of a saving bank to the area for which the public body is responsible." As they were firmly rooted in the local economy, they did not compete with each other; "savings banks in a county or region had reason to consider each other more as peers and colleagues than as competitors." Fifth, they "were part of dense and closely cooperating networks of legally independent institutions that constituted a special banking group." Germans use the term "Verbünde for these dense networks, a term, Bübül, Schmidt, and Schüwer point out, that is hard to translate into English, "since such networks of banks do not exist in Anglo-Saxon countries" (3).

Cooperative banks, the third pillar, were also banks that adhered to the regional principle and were part of dense networks. Their "mandate was to support economic undertakings of their clients and to be cost-covering and profitable businesses. Cooperative banks were organized almost like clubs wherein the owners and providers of equity were not called shareholders but members. The difference between shareholders and cooperative bank members is that the latter could not "sell their shares if they wanted to exit, at some market price, but only hand

them back to the cooperative and in return get back what they had originally paid for them plus their part of the cooperatives accumulated profits.” Accordingly, they could not “benefit from policies that would increase the value of their shares because they could not sell their shares at higher prices” (Bübül et al, 3).

In the 1980s British and American banks and their European partners pushed, as they did in the first pillar, the private commercial banks, to “modernize” the other two. “Strongly opposed to publicly owned banks” EU banking bureaucrats, who were educated in US financialization, joined in, since they thought public savings and cooperative banks old fashioned and outdated, because they did not conform to the “model” of how a good modern bank should be structured and operated.

Reform occurred in Belgium, where savings and cooperative banks essentially disappeared, in the UK where public savings banks (TSB) were sold to Lloyds Banking Group, and several cooperative banks, the so-called building societies, sold to large private banks; in the Netherlands savings banks disappeared and independent cooperative banks were amalgamated into one big national bank (Rabobank); in Sweden the former local savings banks were converted into joint stock corporations in the 1990s and most consolidated into a single national savings bank (Swedbank); in Spain local savings banks, the *cajas*, were privatized and localization abolished. They were permitted to provide a broad range of financial services in all parts of the country, becoming universal banks, which invested heavily in real estate loans, with the approval of pre-financial crisis reformers who believed that regional banks could not compete with other banks operating with large branch networks.

Only in Germany did the other two pillars of banking (423 savings banks and 1,116 cooperative banks) remain a “special case in which no substantial changes [occurred] during the last decades” (Bübül et al, 3). The savings banks have remained local and public and cooperative banks have not become essentially profit oriented institutions seeking to enhance shareholder value; nor has either been turned into centrally located stock-exchange listed corporations. Since each sector had a system of joint and several liability even before the financial crisis began, no individual member bank was allowed when it came to go bust. They came through the crisis with barely a scratch and, their spokesmen argue, their business model, working for the public or mutual good rather than for shareholders, has proved to be well-suited to the mixture of households and small companies (known as the *Mittelstand*) that they serve (Gerada & Netessine,1).

This statement is borne out by their lending record since 2007. Private German commercial banks reduced their medium- and long-term lending to companies and households between 2007 and 2012 in favor of short-term loans, while the German savings and cooperative banks did the reverse. The savings banks and cooperative banks currently provide about two-thirds of all lending to *Mittelstand* companies and 43% of lending to all companies and households.

Most people now agree that “the amazing resilience of the German economy” can be attributed to its reliance on the small to medium size enterprises of *Mittelstand* companies: Seventy percent of Germans are employed by them in the private sector. Inasmuch as private and cooperative banks have financed these flourishing *Mittelstand* firms, judgments about these two pillars of German banking have changed from those of the pre-financial crisis era. Petra Dünhaupt notes that locally rooted banks “compared to private commercial banks,” performed well before and after the crises, (18) and that the modern view that “capital markets, in which banks are large, private, purely shareholder-oriented and exchange-listed corporations has been severely discredited by experience from the recent financial crisis.”

(19). The best business model, she writes, is “being firmly rooted in the local economy and aspiring to strike a balance between the need to make a profit and the aim of serving members and clients, and the appropriate institutional structure is being embedded in a decentralized and dense network of affiliated financial and non-financial institutions” (19).

My own experience illustrates how the second pillar of the banking system worked locally during the financial crisis. In 2004 I bought a ruined house in Görlitz, Germany, the price of which was set by an engineer in a city-affiliated organization, not the market. The purchase agreement stipulated that the new owner had to renovate the building within five years or it could be reclaimed from them by the city at original cost. The construction office in City Hall stipulated that I had to apply to them for the construction permit, and choose the construction foreman from a list of experienced people approved by it. Needing funds, I asked the local branch of Deutsche Bank for 100,000 Euro to help carry out renovations. They rejected the request out of hand; when asked why, they replied, quite arrogantly, that they did not provide explanations. The construction foreman on his own contacted the local savings bank (Sparkasse) and carried out the negotiations with my approval; the bank authorized a 20-year loan at a very low fixed-interest rate. But with stipulations: I had to deposit \$100,000 my Sparkasse account, and pay it out fully on renovation before the authorized Sparkasse renovation loan money could be touched. The bank officers supervised the process. My \$100,000 spent on renovation, the Sparkasse released their renovation money in three installments, paid after their officers verified that the work at each stage had been completed.

The construction foremen noted that the Historic Buildings Preservation Authority (*Denkmalschutzamt*), headquartered in Dresden with a branch in Görlitz, subsidized the restoration of historic buildings' exterior shells; he did the paperwork, the agency approved (thirty percent of costs estimated at 100,000 Euro). The subsidy came from three sources: the city, state (Saxony) and federal budgets – a good example of dense networking.

The city finance office also permitted the VAT paid on renovation labor and materials, starting in 2004, to be deducted from the turn-over tax because I had declared my intent from the beginning to open a bed and breakfast on the site, even though the B&B was not actually opened until 2008. But The Finance Office, to stop speculation, specified that if I sold the property within ten years (dated from 2008), I would have to pay back all the subsidies and tax concessions I had received and pay a very hefty capital gains tax if the selling price warranted it (unless the new owner took over the B&B and ran it). I had to be good citizen, not a foreign speculator. None of these transactions had to do with interest rates set by capital markets, with financialization, or with returns to stockholders in a privatized bank. They did have to do with policies followed by public savings banks and regional nonfinancial agencies that supported local enterprise and city improvement.

Discussions about the distribution of incomes in America and Germany should, therefore, depend as much on how they are embedded in the social and institutional financial-banking systems of each country as on income amounts or the extent of the gap between the top one and bottom ninety-nine percent. Two examples can be used to illustrate this point. One concerns the group composition of the top twenty firms in each country, ranked by revenues (2012).

USA

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

(Source: C Stahl (2013) "Corporate Responsibility in US & German Firms," 59)

Germany

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

(Source: *Ibid*, 61)

Some firms on each list are classifiable under the same rubric, e.g., retail giants (in the US Wal-Mart and McKesson; in Germany the Aldi and Edeka Groups). Others are famous oil and energy firms, mostly on the US list. Whereas few of the firms on the US list were famous before WWII (Ford, GM, GE), such firms dominate the list of the German top twenty, many of them prominent even before the First World War (Deutsche Post, Robert Bosch, Daimler, BASF, Thyssen Krupp, Bayer, and Deutsch Bahn). From a financialization perspective, the big difference is that among the top twenty US firms there are many drivers of financialization (Berkshire-Hathaway, Fannie Mae, Bank of America, JP Morgan Chase Co, Citi-Group, and GE Financial), or US 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 creation, although many well-known German firms went public when the era of financialization began.

The second example is about firm governance and how it affects income distribution. Under director primacy US CEOs set their own salaries using financialization instruments liberally; in large German firms, in the system of co-determination, supervisory boards, which are fifty percent elected by firm employees (usually members of unions) and fifty percent by stockholders, set the salaries of management (the Vorstand). One might think the presence of employees' representatives on supervisory boards would push director salaries down, but this has not especially been the case. Apparently, in periods of prosperity, financialization of top salaries has occurred in large German joint stock corporations, because supervisory boards

have been generous in their granting of stock options, bonus provisions, and high salaries to their managing directors, to the point that *Vorstand* incomes in these big German firms have started to track those that financialization brought to US executives. (Dünhaupt, 2011) Nonetheless, institutionally, co-determination potentially limits upward movement of management salaries in German big business that financialization brings, which the American system of director primacy firm governance does not. More importantly co-determination limits the size of the gap between employees and executives in big firms. The most famous example is in the automobile industry (12% of the Germany economy) where German firms pay their workers twice as much as American. Sathbh Walsche notes, for instance, that VW workers get paid \$67 an hour in Germany but make under \$20 in the VW plant in Tennessee. ([Sathbh Walsche the Guardian.com](#), Wednesday 19 February 2014, also see, Hargreaves, 2014) VW is not an exception among German firms. In Germany the employee elected works councils in the large firms, which have been around as have most of the firms on the German list since the co-determination regime began in the early 1950s, routinely negotiate wages and bonuses with management.

Still, the greatest difference in embeddedness, with respect to financialization, occurs at the level of the small and medium firms. In the US, start-up firms are an integral part of American folklore, for every start-up dreams of one day going public, like Microsoft or Facebook, and turning their founders into billionaires. No matter if the firm's business incomes do not produce the earnings of top executives, the system of financialization that Wall Street presents, permits billionaires to happen. "Too much profit orientation, too much financial sophistication, too much profit pressure emanating from capital markets" occurs (Dünhaupt, 2011, 19).

Few of the financial metrics (about a firm's stock price, about its achieving financial "expectations," or the firm going public) that business television journalists discuss with the financial experts they constantly interview on the evening business news have anything much to say to the German *Mittelstand* and the banking systems they use. German *Mittelstand* firms remain largely unincorporated; no stock option pay regime possible there. German SMEs are primarily self-financed out of earnings, or regionally through traditional bank loans from savings and cooperative banks.

Nor are family-owned German businesses victims of leveraged buyouts by private equity companies, which are integral to American financialization. The conversion rate of these quite successful German SME firms into public owned companies is very low; in Germany from 2000 to 2007 only one percent of successor arrangements for founder or family-run firms involved private equity buyouts. (Schmohl, 2009, 4) Takeovers by foreign private equity firms are in fact particularly disliked. "Although many family businesses are going through generational changes," Josh Kosman writes, "...few have sold out to foreign Private Equity firms" (Kosman, 2010, 168). Germany ranks lowest among the European countries with regard to foreign private equity buyouts.

Not much evidence about SME financialization can be garnered indirectly either, through, for example, information about a family-owned firm's adoption of professional management. Germany's SMEs try to avoid the Buddenbrooks-effect (having incompetent offspring take over from competent founders) by hiring professional managers. But the firms do not seem to adopt the outlook of a professional management caste like that in America. Instead, they usually operate under hybrid management (family+professional managers) from which they profit from management expertise while the firm's metrics are set by owner-families not

professional managers, who in the US under financialization engage in short-term profit maximization and cost cutting at the expense of the workforce.

Rather, the *Mittelstand* firms strive for sustainability within a valued community. The factors supporting this goal become their all-important metrics and they are, like the pillars of banking on which they rely, primarily local and regional not national. Because German SMEs seek sustainability, they make implicit life-time commitments to their employees that they carefully recruit and train, relying on local networks. SMEs consistently consider labor a fixed asset, a human capital, and reject the notion that it is a variable cost that needs to be minimized. Since manufacturers know that sustainability demands staying abreast if not ahead in technology, the firms invest five percent of their revenues into research and development (Vernohr and Meyer, 2007, 29). They also exploit available scientific knowledge and knowhow by working with people in local universities and polytechnics (*Fachhochschulen*), research institutes, and in special places like the Fraunhofer Institutes set up throughout Germany, to facilitate the transfer of scientific research into innovative products and services mainly in *Mittelstand* firms. To promote worldwide success in niche services and manufacturing, on which they concentrate, German SMEs devote much time to cultivating customer relations, owners often making repeated foreign trips in order to establish and maintain personal contacts with customers. Through their business activities German *Mittelstand* firms make high incomes for their owners, but they do it through business activity, not financialization.

Educational systems

I have been publishing articles and books about the comparative development of engineering and business education in Germany, France, the UK, the US, and Japan for over thirty-five years [the first, in 1977, "Industrialisierung und Erziehungssystem in Frankreich und Deutschland vor dem 1. Weltkrieg." *Historische Zeitschrift*, 222, 265-96; the most recent, in 2011, "Reform of Finance Education in US Business Schools," *Real World Economic Review* 58 (December), 95-112.] Since it has been America's great contribution to education, it should be no surprise that the discussion has focused to a large extent on US contributions to business and management education and how they reflect in and foster the transformation in national and international business. Already in 1949-50, 617 US institutions offered courses in business and commerce to 370,000 undergraduates, almost twice as many students as those studying engineering. But the most impressive innovation has come in graduate education – with the big business schools, Harvard, Wharton, Carnegie-Mellon, Chicago, MIT, Georgia Tech, UCLA, Stanford, and a few others leading the pack -- 4,924 MBAs graduated in 1960, 23,400 in 1970, 70,000 in 1980s in a continuous expansion up a steep curve (Locke, 1996, 28). Business school deans and faculties were not shy about cultivating relations with CEOs or about adapting their programs to their needs. Locke and Spender, leaning on Rakesh Khurana's solid study, note how they have also done so in the era of financialization:

"MBAs increasingly found jobs in the banks, hedge funds, and investment houses of the expanding sector. Khurana, cites a survey of first jobs for graduating Harvard Business School 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)" (Locke & Spender, 135).

The system's capacity to expand internationally depended on the receptivity of host countries. In the UK, although without a business school tradition, the MBA idea took hold, first in London and Manchester, and then in a spate of business school creations primarily in the 1980s. In France, the schools of commerce also greeted the American study programs willingly, and the newly established (1968) French Management Education Foundation (FNEGE) developed a program to send hundreds of fledgling French management professors to American and Canadian business schools to imbibe US academic management science. But in two countries the US MBA business school model made little headway: Japan and Germany.

With this fact in mind, I did a little survey in 1983 to try to clarify the basis of job recruitment in the United Kingdom, France, and Germany. I selected a prestige US management consultancy, the Boston Consulting Group, which did major business in these nations. I visited BCG head offices in each country and asked the people responsible for hiring consultants to work for them what educational backgrounds their recruits had. In France the answer was quick and clear; they only hired people from the *grandes écoles* (*Ecole Polytechnique, Ponts et Chaussées, Mines, and ENA*), since graduates from these schools would be running the firms using the consultancy. In London, the answer was a bit perverse. BCG was looking for "high flyers", very bright people who could talk to top executives in top firms about big problems, like strategy. The firm recruited its consultants from the best US business schools and Oxbridge, where at the time there were no business schools. Thus, despite the existence of London and Manchester business schools for almost twenty years, in 1983 the firm preferred to get people who had studied the classics (Greats) or PPE (Politics, Philosophy, and Economics) in Oxbridge. This mode of selection would probably upset people in elite business schools, because they think something taught by them is more useful to consultants than the classics, but it made sense as far as BCG clientele were concerned. A BCG consultant would deal with a social type (public school, Oxbridge) whose favor they would have to curry to succeed in their consultancy. To gain their confidence, to appear intelligent and capable, a consultant had to be liked by their clients. As in France, a person from the wrong school, with the wrong accent, would have an uphill struggle not because he/she were incapable or ignorant but because perceptions of ability and intelligence are socially shaped. Oxbridge fit that bill. (Interview results given at various places in Locke, 1989.)

At the Munich office of BCG the recruiter, Dr Struve, answered my questions with a lament. "In Germany there are no national prestige schools, like the French *grandes écoles*, the elite US business schools, or Oxford and Cambridge, that can screen out the best candidates for us. We have to do the screenings and interviewing ourselves from a large pool of candidates that have attended lots of good schools; we select not by school attended but by disciplines studied." When pressed on the issue, Dr Struve asserted that the favored candidates were engineers or economics-engineers (Wi-Ing., a study program composed fifty percent of engineering and fifty percent of business economics courses). This absence of nationally elite schools made recruitment more work for Dr Struve. When big German commercial banks in the 1990s decided to adopt the US-UK investment banking model, they even had trouble recruiting in Germany on the basis of disciplines studied. In Germany faculties of economics (*BWL*) studied finance, but finance professors did not have the contacts in investment

banking that prestige US business school finance professors (at Wharton, Harvard, MIT, Chicago, Columbia, Stanford, and others) had built up earlier in the era of US financialization. German commercial banks decided to develop the required expertise through acquisition. Deutsche Bank turned to the UK and the US to recruit staff well versed in the ways of capital markets, and it bought Morgan Grenfell, the British merchant bank in 1989 and Bankers Trust, the US specialist in hedge funds, in 1999. Dresdner Bank acquired UK-based Kleinwort Benson in 1995 and US-based Wasserstein Parella in 2000, attempting to expand into the global big leagues of underwriting, sales and trading, and merger advice. Deutsche Bank established its investment branch in London. To satisfy their educational needs, they drew substantially on those educated in America because of German educational deficiencies.

The key point about German education, however, is if ill equipped for financialization of the big German commercial banks, it was well equipped to serve the educational needs of the *Mittelstand*. This is true because Germans have a different conception of education from people in Britain and America. I explained an important aspect of the difference In "Reassessing the Basis of Corporate Performance," (*Real-World Economics Review*, 2013) I wrote:

"Ian Glover notes that 'In Anglophone countries, two cultures, the arts and sciences are recognized.' In the two cultures engineering is placed in an inferior place within the science culture, and UK scientists looked down on engineering as an inferior subject for the less brilliant and gifted. Glover went on to note that in [Germany] rather than two cultures there are three: '*Kunst* (like the arts), *Wissenschaft* (similar to science) and *Technik* (the many engineering and other making and doing subjects, representing practical knowledge (*Können*),' including scientific knowledge (*Wissen*). (Glover, 2013, 9) In Germany a great chain of practical (tacit) education (*Können*), the art of practical work, topped off with knowledge (*Wissen*) gained primarily in technical *Hochschulen*, combined, in education and workplace, to define German engineering as this third culture of *Technik*. The German engineering society [Verein Deutscher Ingenieure (VDI)] has consistently pitched a large tent, including in its membership craftsmen, machinists as well as university educated engineers. They stood and stand as equal participants through their skills and knowledge."

Within the world of *Technik*, the justly famous German apprenticeship system forms the practical technical and commercial educational base. German secondary school students can and do enter into an apprenticeship, after grade 10, and work in an organized program four days a week on some approved occupation (chimney sweeping, bookkeeping, banking, metal working, machine operating – there are over 400 options), while still attending secondary school courses (e.g. in English, German, mathematics) two days a week, before they end the program, after three years, if successful, with an apprentice certificate in their specialty (*Fach*). This practical, tacit education is primarily carried out in *Mittelstand* firms. It is local, and can be continued up to the master craftsman's level (*Meisterbrief*), a qualification that is highly respected in the German work world. Often first line supervisors in German factories have this qualification. A network of local institutions, in particular, the *quasi* public Chambers of Commerce and Industry are charged with administering the programs, and the employee-elected works councils in firms where pupils are apprenticed monitor the effectiveness of the training.

The apprenticeship certificate is not an educational dead end. There is the possibility of more extensive practical learning through the preparation of a *Meisterbrief* or to explore the more scientific dimension of *Technik* by enrolling in local technical and commercial schools. These schools have their roots in practical education, too, specifically because throughout most of the 20th century people had to have completed an apprenticeship to gain entry into them. Their students study for three years, alternating coursework with work stints in praxis. Despite the practicality of this education, the students through their faculties are also exposed to science, since the schools' teachers have to have degrees in subjects with a scientific input (explicit learning) from German universities (*Hochschulen*). Subuniversity graduates (originally Grad-Ing and/or Grad-Kauf, later Dipl-Ing FH and/or Dipl-Kauf FH) from these schools (now called *Fachhochschulen*) have consistently been highly sought after by German industrial firms; they rise to the highest positions in German industrial management.

Sixty percent of secondary school pupils participate in apprenticeship programs mostly in *Mittelstand* firms; among the forty percent who do not, ten leave school with poor qualifications while the other thirty percent finish their A levels (*Abitur*) and go on to study at university in great numbers. But unlike the American, British, and French elite who leave home for prestige schools and the national and/or international employment scene, the Germans, concentrating on subjects instead of the reputation of schools, stream into the excellent regional universities that are the strength of their university system, with every chance of remaining regionally oriented after their studies are completed, serving the educational needs of the *Mittelstand*.

If German *Mittelstand* incomes thrive, from the point of view of social justice it is a much better outcome than the exorbitant incomes the top one percent of Americans enjoy through financialization. Clearly German firms do a lot for their communities and are respected for it. This cooperative sense even extends to big firms where top executives do profit from financialization. While in the 1980s director primacy governance in league with the Republican politicians wrecked the social pact with workers in America, German conservative parties (CDU & CSU) never turned their back on co-determination; Helmut Kohl and Angela Merkel remain faithful to it. (Locke & Spender, 80) But more importantly, so do the employer associations; the German Employer Confederation and the Employer Association for the Chemical Industry (Bundesarbeitgeberverband Chemie), for example, opposed reforms at the end of the 20th century designed to advance the interest of firms and managers, which would have gutted co-determination. (Stahl, 60, Werder & Grundei, 101)

Conclusion

There are clear signs about what needs to be done to diminish the effect that financialization has on income distribution. It is obvious that the solution to the problem of excessive and wildly mal-distributed incomes is not to set up ethics courses for MBA students at Harvard, London, the Chicago Business School, and elsewhere (Locke, 2011b). Solutions require the adoption of new public policies and legal-institutional change. They involve politics and are about grasping power. Nor should political control be sought primarily in underdeveloped and/or developing countries, where financialization wreaks havoc. The West is not driven by some financialization monolith; there are strong advanced economies, as the German example shows, and a political base, even within the business community, that is ready to oppose this juggernaut. To choose is simple: If people want to keep out undesirables from their community why just pass anti-immigrant or vagrancy laws; they need also to stop rich

financial interlopers in private equity firms from buying local firms and using bankruptcy statutes to deprive employees of their pension and benefit plans. They also need, like the Germans do, to give employee representatives on supervisory boards a voice in setting the salaries of top management and in firm governance, so that they can resist acquisitions and takeovers. It won't be easy; witness American workers' (under intense pressure from Republicans and the business community) recent rejection of the union at Volkswagen's plant in Tennessee, which spoiled the company's attempt to introduce a works council in the plant (Volkswagen is fully unionized with works councils included in its governance everywhere in its worldwide operations, except Tennessee).

Something needs to be done to counter the takeover of educational institutions by private interests. In the nineteenth century the Morrill Act (1859) set up land grant colleges throughout America "to promote the liberal and practical education of the industrial classes in the several pursuits and professions in life." They were, like the technical and commercial institutes being established in Germany, meant to promote the public good through agricultural and engineering education. The business school movement in the U.S. and its extension overseas has taken a different route. Already in his 1918 book *The Higher Learning in America: A Memorandum on the Conduct of Universities by Business Men*, Thorstein Veblen provided a critical perspective on the role of the schools of commerce (today called business schools) within the American university and, by consequence, their effect on the society as a whole. He asserted that "the college of commerce, if it is to live and thrive, may be counted on to divert a much larger body of funds from legitimate university uses, and to create more of a bias hostile to scholarly and scientific work in the academic body, than the mere numerical showing of its staff would suggest" (Veblen, 1918, 157). Furthermore, he wrote about the consequences that a "habitual pursuit of business" has on the ideals, aims and methods of the scholars and schools devoted to "the higher learning". Put simply, "The consequences are plain. Business proficiency is put in the place of learning" (Veblen, 1918, 142 in Robert Kemp, 2011). He might have added business proficiency is put at the service of the perpetuation of a moneyed elite, for these schools, where tuitions range up to \$100,000 a year, are private welfare clubs for the upper classes, supported with lavish endowments from businessmen for the schools' academic chairs and the schools themselves. Financially hijacking public institutions to promote private greed is not philanthropy for the public good (Locke, 2012, 110-11). It is part of financialization and should be appropriately dealt with in the tax codes. Germans call the purpose of higher education *Wissenschaft* (scientific investigation). Follow them. Need more be said?

References

- Bainbridge, S. (2006). "Director Primacy and Shareholder Disempowerment," *Harvard Law Review*, Vol. 119, UCLA School of Law – Econ Research Paper No. 05-25 – available at SSRN <http://ssrn.com/abstract-808584>.
- Ball, R & Appelbaum, E. (2013). "The Impact of Financialization on Management and Employment Outcomes." *Upjohn Institute Working Paper*, 13-191. Kalamazoo, Mi: W. E. Upjohn Institute for Employment Research.
- Bátiz-Lazo, B., Müller, K., & Locke, R. R. (2008). Transferring Rhineland Capitalism to the Polish-German Border. *International Journal of Bank Marketing* 26(2), 76-98.
- Bülbül, D., Schmidt, R. & Schüwer, U. (2013). *Savings Banks and Cooperative Banks in Europe*. House of Finance Goethe University, Frankfurt am Main, White Paper Series No. 5 (August 20).

Chandler, A. D., Jr. (1977). *The Visible Hand: The Managerial Revolution in American Business* (Cambridge, Mass.: Harvard University Press).

Dünhaupt, P. (2011). The Impact of Financialization on Income Distribution in the USA and Germany. Retrieve on line at Epstein, G. (ed) (2005). *Financialization and the World Economy*. Cheltenham: Edward Elgar.

Dünhaupt, P. (2012). "Financialization and the rentier income share – evidence from the USA and Germany." *International Review of Applied Economics*. Taylor & Francis Journals, vol. 26(4), pages 465-487, June.

Glover, I. (2013). "Bleak House: Pessimism about Management, Responsibility, and Society in the Early Twenty-First Century." *Work, Employment, and Society*, April. 1-10.

Girota, K. & Netessine, S. (2013). "Extreme Focus and the Success of Germany's Mittelstand." *The Harvard Business Review*, Feb. 12.)

Kemp, R. (2011). "The Business School in the Corporation of Higher Learning in America." *Journal of Pedagogy*. Vol 2, No 2, 283-94.

Hargreaves, D. (2014), "Can We Close the Pay Gap?" *NY Times. Opinionator*. March 30. p. SR3.

Khurana, R. (2007). *From Higher Aims to Hired Hands: The Social Transformation of American Business Schools and the Unfulfilled Promise of Management as a Profession*. Princeton, NJ: Princeton University Press.

Kosman, J. (2009). *The Buyout of America: How Private Equity Will Cause the Next Great Credit Crisis*. New York: Portfolio.

Locke, R. R. (1989). *Management and Higher Education Since 1940*. Cambridge: Cambridge UP

Locke, R. R. (2008). "Comparing the German and American Systems." Roundtable on Business Education. A Consideration of Rakesh Khurana's *From Higher Aims to Hired Hands*. *Business History Review*, 82(2), 336-342.

Locke, R. R. (2011a) *Management from Hell: How Financial Investor Logic Hijacked Firm Governance*. (Paris: Boostzone Editions. An ebook).

Locke, R. R. (2011b). "Reform of Finance Education in U.S. Business Schools: An historian's view," *Real-World Economics Review*, issue no. 58, 12 December, 95-112.

Locke, R. R. (2012). "Reassessing the Basis of Economics: From Adam Smith to Carl von Clausewitz." *Real-World Economics Review*, issue no. 61, 6 September, 100-114.

Locke, R. R. & Spender, J-C. (2011). *Confronting Managerialism: How The Business Elite and Their Schools Threw Our Lives Out of Balance*. London: Zed Books, New York: MacMillan Palgrave.

Mitchell, L. E. (2008). *The Speculation Economy: How Finance Triumphed Over Industry*. San Francisco CA: Berrett-Koehler Publishers.

Lütz, S.(2000). "From Managed to Market Capitalism? German Finance in Transition." Discussion Paper 00/2, Max Planck Institute. Cologne.

Schmohl, J. (2009). "Entrepreneurial Exit Management -- Key Success Factors of the Private Equity Buyout Options." PhD Dissertation No. 3792. University of St Gallen, Graduate School of Business Administration (HSG) (19 October).

Stahl, C. (2013). "Corporate Social Responsibility in US and German Firms." Master's Thesis. Graduate School of Business, University of Grenoble.

Veblen, T. (1918). *The Higher Learning in America: A Memorandum on the Conduct of Universities by Business Men*, Retrievable at <http://etext.virginia.edu/toc/modeng/public/VebHigh.html>.

Vernohr, B., & Meyer, K. E. (2007). The German Miracle Keeps Running: How Germany's Hidden Champions Stay Ahead in the Global Economy. Paper, Institute of Management, Berlin School of Economics.

Werder, A. & Grundei, J. (2001). "Generally Accepted Management Principles (GAMP) – functions, first proposals, and acceptance among German top managers. *Empirical Research-Based and Theory-Building Papers*. Technical University Berlin. Vol. 9 No. 2 101-109.

Author contact: lockerobert3@aol.com

SUGGESTED CITATION:

Robert R. Locke, "Financialization, income distribution, and social justice: recent German and American experience", *real-world economics review*, issue no. 68, 21 August 2014, pp.74-89, <http://www.paecon.net/PAERreview/issue68/Locke68.pdf>

You may post and read comments on this paper at <http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

Recovering Adam Smith's ethical economics

Thomas R. Wells [University of Groningen, Netherlands]

Copyright: Thomas R. Wells, 2014
You may post comments on this paper at
<http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

While some men are born small and some achieve smallness, it is clear that Adam Smith has had much smallness thrust upon him (Sen 2010, 54).

Adam Smith is famous for founding economics as an independent field of study by synthesising and systemizing classical economics in *The Wealth of Nations* (1776). But he was also a significant moral philosopher in his own right whose *Theory of Moral Sentiments* (1759) was respected by his contemporaries, including Immanuel Kant and his close friend David Hume. In this brief essay I want to right a wrong by showing how deeply Smith's ethics was integrated into his approach to economics. Smith saw economics as a branch of moral philosophy (Griswold 1999, chap. 1), and he saw capitalism as an ethical project whose success required political commitment to justice and freedom, not merely an understanding of economic logistics.

These days Smith is best known as an economist, and specifically as the defender of the famous "invisible hand" of free-market economics, wherein the self-interested actions of private individuals, mediated through free markets, generate results that are good for society as a whole. The market-system comprehends the true level of demand for any good and provides the appropriate incentives—profits—for producers to adjust their output to match. No external intervention or guidance is necessary. A great deal of contemporary (neo-classical) economics can be understood in terms of translating Smith's invisible hand metaphor into a systematic theoretical form, with a particular emphasis on the economic efficiency of perfectly competitive markets.

However the popular view of Smith among economists that has resulted from this emphasis is twice distorted. Firstly, it is based on the narrow foundations of a few select quotations from *The Wealth of Nations* (WN) that are taken in isolation as summing up his work (Smith only mentions the "all important" invisible hand once); and secondly, these quotations have been analyzed in a particularly narrow way. Both selection and interpretation have been driven by contemporary mainstream economists' interest in justifying orthodox economic methodology and their peculiar (Mandevillian not Smithian) assumption of the selfish utility-maximising *homo economicus*. The Chicago School economist George Stigler once famously declaimed, "The Wealth of Nations is a stupendous palace erected upon the granite of self-interest" (Stigler 1975, 237). What such Whig "historians" have achieved is the diminution of Smith's economics to those bits which can be claimed to be early (and flawed) fore-runners of contemporary economic concepts and techniques.¹

But anyone who cares to read Smith's *Wealth of Nations* for themselves will find an economics discussed and justified in explicitly moral terms, in which markets, and the division of labour they allow, are shown to both depend upon and produce not only *prosperity* but also

¹ See, for one example among many, the Whiggish mistreatment of Smith's trade theory in contemporary history of economics textbooks analysed by Reinhard Schumacher (Schumacher 2012).

justice and *freedom*, particularly for the poor. With those concerns in mind, it should not be surprising that Smith was a staunch and vehement critic of those particularly grotesque sins associated with early capitalism: European empires and the slave trade.²

Smith's defence of capitalism (or, in his terminology, "commercial society") is unambiguous but qualified. There is no inconsistency here. Smith's commitment to a realistic liberalism led him to endorse commercial society over any previous socio-economic system as a social order in which the most people possible could live decent lives. But he was not the blind zealot for the market he is now sometimes portrayed as. Smith was acutely aware of the possible ethical shortcomings of commercial society and, for example, carefully read and responded to Rousseau's powerful critiques of its materialism, inequality, and inauthenticity (Rasmussen 2008; Hanley 2008). While the structural features of commercial society set the terms of its main opportunities and challenges, they did not determine the outcome. Commercial society was for Smith an ethical project whose greatest potential benefits had to be struggled for, and which could and should be much better than it was.

The Enlightenment concern for perfecting social order was both the background to Smith's thinking and a goal Smith eschewed. As Rousseau put it in *The Social Contract*,

The problem is to find a form of association which will defend and protect with the whole common force the person and goods of each associate, and in which each, while uniting himself with all, may still obey himself alone, and remain as free as before ([1762] 2008, sec. I.6).

While Rousseau sought a perfect and absolute solution to the problem through his famous social contract, Smith argued that, under conditions of freedom and justice, society could endogenously produce a decent social order for co-ordinating moral and economic conduct without centralised direction or coercive moral policing by religious or secular authorities. In this sense his project can be seen as a working out of Locke's liberal political philosophy at the institutional level. Smith was not interested in what a perfect society might look like, but rather with understanding the world as it was and how it might be improved, i.e., with real world economics. So instead of analysing the requirements of a perfectly just society he analysed the socio-economic order of the new commercial society then coming into being, characterized by an enormously increased division of labour, dependence on strangers, formal property rights, and individual mobility. And he saw that commercial society had enormous potential for enhancing general prosperity, justice, and freedom.

Prosperity

Smith analysed the wealth of a nation as the ability of its ordinary citizens to command goods to satisfy their wants, i.e., not only the total wealth but also its *distribution*. Smith noted that a European peasant was now materially better off than many African kings (WN I.i.11), but he attributed this not to any innate European superiority (as all too many 19th century political economists went on to do) but to changes in political economy. The recent increase in the wealth of certain nations was due to the increasing role of markets in their economies, which

² With regard to the latter, see the debate in these pages between Marvin T. Brown, Bruce Elmslie and myself in 2010 (Brown 2010a; Elmslie 2010; Wells 2010; Brown 2010b).

made possible and rewarded the technical innovations and efficient organisation of labour that dramatically increase productivity.

That benefits the ordinary citizens of a country—i.e. the working poor—in two ways. Firstly, when producers compete fairly and freely with each other to supply the public with cheaper (and better) products there is a natural tendency for the market price to fall towards the actual cost of production, and for the costs of production themselves to fall, meaning cheaper products for consumers (and less profits for producers). This aspect is central to contemporary mainstream defences of the market. But Smith also noted that the rise in labour productivity meant that wage labourers (the bulk of the population) could exchange their labour for a greater command of those goods. Thus, Smith praised the expansion of markets for their role in increasing the purchasing power of ordinary citizens and thereby the *real wealth* of a nation. A concern for equitable distribution was constitutive of his understanding of prosperity and the subject-matter of economics. As he noted:

No society can surely be flourishing and happy, of which the far greater part of the members are poor and miserable. It is but equity, besides, that they who feed, cloath and lodge the whole body of the people, should have such a share of the produce of their own labour as to be themselves tolerably well fed, cloathed and lodged (WN I.viii.36).

Justice

Smith's commitment to "equity" for the working class was behind the vehemence of his opposition to mercantilist ("business economics") arguments for policies that would protect or promote the profits of producers and intermediaries. Smith saw such pro-business arguments—which arguably persist as the core of neoliberalism (Harvey 2007)—whether for direct subsidies or competition-restricting regulations, as an intellectually bankrupt and often morally corrupt rhetorical veil for what were actually "taxes" upon the poor (what we now call "rents").³ Such taxes are unjust and outrageous because they violate fair play both in the deceptive rhetoric by which they are advanced and by harming the interests of one group in society (generally, the poor and voiceless) to further the interests of another (unsurprisingly, the rich and politically connected). Smith explicitly moralised the point,

To hurt in any degree the interest of any one order of citizens, for no other purpose but to promote that of some other, is evidently contrary to that justice and equality of treatment which the sovereign owes to all the different orders of his subjects (WN IV.viii.30).

Justice was thus central to Smith's critique of the crony capitalism of his time, and to his alternative proposal of a "system of natural liberty" characterised both by a level playing field (the responsibility of political institutions) and a commitment to "fair play" (the moral responsibility of economic actors). The quotation above is often taken to indicate Smith's rejection of the interests of the poor by ruling out the kind of redistributive policies found in a modern welfare state as akin to a referee changing the results of a game to favour one "team"

³ See for example Smith's scathing criticism of the ban on wool exports to promote the interests of English textile manufacturers, which was imposed by particularly onerous methods (WN IV.viii).

over another. Yet that misses Smith's commitment to *procedural fairness*, which introduces a concern that the rules of the game—the institutional arrangements that decide who should get what share of the gains of economic activity—should themselves be fair. If a country's economy creates great wealth but the share going to the workers versus the owners of capital is kept artificially low by unfair institutions—such as restrictions on workers' ability to bargain (WN I.viii.13)—that is a gross injustice which keeps the country less prosperous than it ought to be. Smith thus appears a more radical critic of the structural origins of economic inequality than many today on the political left. In Smith's time no less than in our own, a political commitment to a free society and a free economy does not imply that we should simply accept our existing socio-economic institutional arrangements (cf Grusky 2012). On the contrary, it implies rigorous scrutiny and reform.

But Smith's moral condemnation of mercantilism has further relevance to today's business-economics dominated policy discussions. According to Smith's diagnosis, the mercantilist system's great success was in nationalising the corporation model of towns in the feudal system, leading to great efficiency gains as the size of the market increased. But in doing so it had also nationalised the “underling” ethics of monopolist tradesmen and manufacturers, who preferred to lobby collectively for self-serving rights and privileges at the political level than to compete on equal terms with others in the market. The “impertinent jealousy of merchants and manufacturers” when coupled to political influence allowed the hijacking of the state's power and authority to promote the interests of a well-connected few in the name of the national interest, such as the extractive economic policies that Smith considered had driven the American colonies to revolt. But this was due not only to straightforward interest group capture but also the ideological capture of the state by the particular—skewed—perspective of merchants and manufacturers. That fostered an invidious political ideology: a zero-sum view of trade as competition rather than cooperation, in which the prosperity of other nations is seen as national defeat. This remains with us today, deeply lodged in the “common-sense” understanding of our politicians, many of their advisers, and self-appointed media pundits. In Smith's day, UK plc competed against France Ltd; now we are all supposed to fear the rise of China Inc.

It should be obvious by now that Smith was no cold heartless utilitarian who put his faith in a ghostly Invisible Hand. But he was a professor of rhetoric as well as moral philosophy, and he was acutely aware of who the likely readers of the *Wealth of Nations* would be. So he supplemented his arguments for the moral priority (even sacredness) of justice with hard-nosed utilitarian arguments about its instrumental role in social order and economic development. For example, when people gain equality before the law and thus security from the predations of the powerful, they have the security they need to make the investments that increase productivity.⁴

⁴ See for example Smith's discussion of the legal protections of tenancy extended to English yeoman farmers, which, together with their rights to political representation, “have perhaps contributed more to the present grandeur of England than all [the] boasted regulations of commerce taken together” (WN III.ii.14). Unsurprisingly the legal property rights of the poor is also an important theme of contemporary development economics, perhaps most notably in the work of Hernando de Soto, though it has unfortunately and unnecessarily become associated with a general neo-liberal programme.

Liberty

Freedom from constraints, freedom from domination, and the freedom of moral autonomy were also central to Smith's economics. Smith is of course most associated with the first of these, also called classical or negative *liberty*, because of his famous endorsement of the "natural system of liberty" in which:

Every man, as long as he does not violate the laws of justice, is left perfectly free to pursue his own interest his own way, and to bring both his industry and capital into competition with those of any other man, or order of men (WN IV.ix.51).

A great deal of WN is concerned with identifying and criticising the artificial and unjustified obstacles placed in the path of ordinary people trying to get on with their own lives and to better their condition. Smith argued forcefully, and in great econometric detail, that England had become richer in spite of and not because of the government's mercantilist policies.

But such mercantilist regulations were wrong not only because they reduced economic efficiency by reducing and distorting competition (the hard-nosed utilitarian argument). They were also wrong because of the insufferable impertinence of a government (or any other body) taking it upon itself to manage people's affairs on their behalf. This was not only a matter of the freedom of great merchants to engage in high international trade. Smith was particularly exercised about the 18th century English laws that deprived wage-labourers, whose only means of subsistence was to sell their labour, of the right to change occupations, negotiate wages, or even move around the country. Of course that produced an inefficient allocation of economic resources: not enough wheel-wrights in one place, too many in another. But even more importantly it disrespected the natural right of ordinary people to make decisions on matters of the greatest consequence to them, which were no-one else's proper business, and about which they were best placed to judge. These policies, justified by chimerical arguments about the public good, reduced and distorted the options available to ordinary people to help themselves and through such artificial helplessness induced real hardship and destitution.

One should note that liberty was a maxim for Smith rather than a dogma (a goal, not a side-constraint as the libertarian Robert Nozick might put it), and he was in favour of government interventions and regulations properly justified by the public interest. As to interventions, while Smith's "laissez-faire economics" proposed taking government out of the business of micro-managing the economy, it also outlined a clear and extensive government responsibility for ensuring the conditions for a flourishing free and just economic system. That included organising (though not necessarily directly providing) public goods that private market actors did not have the necessary credibility, scale, profit-incentives, or long-term perspective to provide, including legal justice, universal education, and security.

Smith's regulatory proposals were directed at preventing systemic failures and some remain highly pertinent. For example, he proposed banking regulations which though "in some respect a violation of natural liberty" upon a few individuals were justified by the government's duty to protect "the security of the whole society" (WN II.ii.94). And he argued for fixing the rate of interest at a relatively low level (just above the prime market rate) in order to prevent imprudent "prodigals" (sub-prime borrowers) and "projectors" (speculators with crazy South

Sea Bubble type schemes) from getting access to credit and thus diverting it from prudent investment and putting the financial system at risk. Unlike supporters of the Efficient Market Hypothesis (and his contemporary Jeremy Bentham (Bentham 1818 [1787]), Smith saw prudence as a personal virtue, more cautious than enterprising, and thought markets were good at teaching it, but unfit to substitute for it. The economists who promoted deregulation of banking and lending in recent times on the grounds that “the market always knows best” may have cited Smith, but they hadn’t read him.

Freedom from domination

Smith also argued that commercial society produced freedom from domination (or “republican freedom” in modern terminology). The feudal system that Smith describes as preceding commercial society (and whose traces could still be seen in his own time in parts of Scotland) was a society characterised by direct relationships of dependence; a world of great landowners with the absolute power of lords over their tenant farmer subjects and retainers. Contemporary communitarians like Alasdair MacIntyre and Charles Taylor may portray such relationships as the basis for an “authentic” human life, or rhapsodise over the “enchanted world” in which such people lived. But Smith cut through the romance of feudalism and analysed them as master-slave relationships that reduced the humanity of all parties.

The appearance of commercial society changed all that for the better. In commercial society informal webs of mutual obligation are transformed into formal consensual relationships between independent agents because these are far more economically productive (feudalism is out-competed). The division of labour mediated by extensive markets replaces closed relationships of direct dependence, in which some must subordinate themselves to the whims of their masters and curry favour to survive, with open networks of inter-dependence spread among the thousands of people involved in producing and bringing to market the most ordinary essentials of life.⁵

On the production side, this liberates workers to sell their labour without having to sell their souls. If people find the working conditions in one employment oppressive they are allowed to take their labour elsewhere. In markets themselves the very fact that people interact as relative strangers, and therefore appeal to each other’s self-interest rather than their benevolence (as beggars must), means that they meet in conditions of relative equality where they must endeavour to persuade others of the qualities of their goods by the gentle arts of persuasion. In this sense, markets economise on love, which is a good thing because, as beggars know all too well, love is scarce.

Moral autonomy

Smith believed that personal autonomy—self-determination—could flourish in commercial society, because its circumstances gave the greatest possible number of people access to the basic requirements for moral self-development. The increased wealth and security that followed a proper administration of justice allowed the mass of ordinary people—not only the aristocratic elite—the leisure to reflect about matters beyond their daily subsistence. Freedom

⁵ See WN I.i.11 for an evocative description of the distributed production of goods in commercial society.

from artificial constraints and domination allowed them to control important aspects of their own lives, from religion to employment, while taking greater responsibility for how they lived. The markets themselves could be schools for certain virtues (the “doux commerce” thesis also associated with Montesquieu (Hirschman 1982)). For example, people who worked for themselves would be more prudent and temperate; people who interacted through markets would be more honest than when trapped in sycophantic relationships with masters (Wells and Graafland 2012). As a result, Smith considered commercial society compatible with the moral autonomy of its ordinary citizens, and believed that such societies would exhibit more moral decency, though less moral greatness, than either classical or contemporary “savage” societies.⁶

Unlike the classical philosophers with their metaphysical elitism, Smith was an enlightenment liberal who firmly believed in the fundamental equality of human beings, and attributed differences in status and achievement far more to the effects of circumstances than to innate qualities. “The difference between the most dissimilar characters, between a philosopher and a common street porter, for example, seems to arise not so much from nature, as from habit, custom, and education” (WN I.2.4). No matter how far you rise, you are still fundamentally equal to others. (Even “philosophers”.) With these liberal commitments in mind it is not surprising that Smith celebrated the possibility for ordinary people to have the leisure, freedom, and education to reflect on their moral lives and decide for themselves who they should be. Nor that he thought the “boring” bourgeois decency they would tend to attain was a social achievement that outweighed the loss of aristocratic-romantic virtues like magnanimity, courage, or “authenticity”.

Conclusion

Smith’s economic analysis was thoroughly entangled with a deeply humanistic ethical perspective. The picture of the real Adam Smith this reveals is of a “true friend” of commerce, supporting the project because of its achievements and its even greater potential, but constructively critical about both the shortcomings of the mercantilist society he lived in and commerce in general. He endorsed commercial society for its tremendous contribution to advancing the prosperity, justice, and freedom of all its members, and most particularly of the poor and powerless in society. But he was no naive ideologue for free markets and profits. He criticised the political machinations and moral character of the very merchants and manufacturers who, he acknowledged, were driving economic development. He not only argued that they should act better, but also proposed institutional measures to restrict their worst proclivities, particularly by getting government out of the business of economic micro-management and thus out of the business of crony capitalism. Though its promise was great, the rise of commercial society meant the loss or sidelining of traditional values and ways of life, and posed new challenges of its own. Its success was not predetermined, but had to be worked for. That is a lesson some modern economists and politicians would do well to relearn.

⁶ The reasons for Smith’s scepticism that moral excellence would thrive in commercial society despite people’s greater opportunities to live an excellent life relate to his recognition of the psychological attractions of the material success that would also become more generally accessible, for “An augmentation of fortune is the means by which the greater part of men propose and wish to better their condition. It is the means the most vulgar and the most obvious” (WN II.3.28).

References

- Bentham, Jeremy. 1818 [1787]. "Defence of Usury: Letter XIII to Dr. Smith, on Projects in Arts, &c." In *Glasgow Edition of the Works and Correspondence of Adam Smith*, 4th ed. Vol. 6. London: Payne and Foss. <http://www.econlib.org/library/Bentham/bnthUs3.html>.
- Brown, Marvin T. 2010a. "Free Enterprise and the Economics of Slavery." *Real-World Economics Review*, no. 52: 28–39.
- . 2010b. "Adam Smith's View of Slaves as Property: A Response to Thomas Wells and Bruce Elmslie." *Real-World Economics Review*, no. 55 (December): 124–25.
- Elmslie, Bruce. 2010. "Did Smithian Economics Promote Slavery?" *Real-World Economics Review*, no. 53 (June): 150–55.
- Griswold, Charles L. 1999. *Adam Smith and the Virtues of Enlightenment*. Cambridge University Press.
- Grusky, David B. 2012. "What to Do About Inequality." *Boston Review*, March 1. <http://www.bostonreview.net/grusky-forum-inequality>.
- Hanley, Ryan. 2008. "Commerce and Corruption: Rousseau's Diagnosis and Adam Smith's Cure." *European Journal of Political Theory* 7 (2): 137–58.
- Harvey, David. 2007. *A Brief History of Neoliberalism*. Oxford University Press.
- Hirschman, Albert O. 1982. "Rival Interpretations of Market Society: Civilizing, Destructive, or Feeble?" *Journal of Economic Literature* 20 (4): 1463–84.
- Rasmussen, Dennis Carl. 2008. *The Problems and Promise of Commercial Society*. Penn State Press.
- Rousseau, Jean-Jacques. 2008 [1762]. *The Social Contract*. Ed. G. D. H. Cole. New York: Cosimo.
- Schumacher, Reinhard. 2012. "Adam Smith's Theory of Absolute Advantage and the Use of Doxography in the History of Economics." *Erasmus Journal for Philosophy and Economics* 5 (2): 54–80.
- Sen, Amartya. 2010. "Adam Smith and the Contemporary World." *Erasmus Journal for Philosophy and Economics* 3 (1): 50–67.
- Smith, Adam. 1975 [1759, 1790]. *The theory of moral sentiments*, eds. D. D. Raphael, and A. L. Macfie. Oxford: Clarendon Press.
- Smith, Adam. 1976 [1776, 1789]. *An inquiry into the nature and causes of the wealth of nations*, eds. R. H. Campbell, A. S. Skinner and W. B. Todd. Oxford: Clarendon Press.
- Stigler, George. 1975. "Smith's Travels on the Ship of State." In *Essays on Adam Smith*, edited by Andrew S. Skinner and Thomas Wilson. Oxford: Clarendon Press.
- Wells, Thomas. 2010. "Adam Smith's Real Views on Slavery: A Reply to Marvin Brown." *Real-World Economics Review*, no. 53 (June): 156–60.
- Wells, Thomas, and J.J. Graafland. 2012. "Adam Smith's Bourgeois Virtues in Competition." *Business Ethics Quarterly*, 22(2), 319-350.

Author contact: t.r.wells@dunelm.org.uk

SUGGESTED CITATION:

Thomas R. Wells, "Recovering Adam Smith's ethical economics", *real-world economics review*, issue no. 68, 21 August 2014, pp. 90-97, <http://www.paecon.net/PAEReview/issue68/Wells68.pdf>

You may post and read comments on this paper at <http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

The human element in the new economics: a 60-year refresh for economic thinking and teaching

Neva Goodwin [Tufts, USA]

Copyright: Neva Goodwin, 2014
You may post comments on this paper at
<http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

Abstract

For more than half a century the discipline of economics has been based on an inadequate and misleading description of human nature. Translated into what students remember, and what has increasingly risen to the top in Anglo-American culture, this description promotes the idea that only selfishness is rational.

Many economists and others have noted that the working out of the “rational economic man” image denies both the social and the moral nature of humankind. At the same time it produces theoretic results, and influences some real-world behaviors, in ways that mimic the initial assumption.

The first four sections of this paper will describe aspects of the old economic theory that are based on the rationality axiom. Sections 5 and 6 will outline some of the findings of the relatively new school of behavioral economics, which is leading a major challenge to the neoclassical economic edifice – the first such challenge that (so far) mainstream economics has been unable effectively to marginalize.

Sections 7 and 8 will build on behavioral economics with attention to some issues in human motives and behavior that should be addressed in a reformulated economic theory. This will include a reconsideration of rationality, as well as discussions of the social and ethical contexts for economic behavior.

The last two sections of the paper will propose some steps toward the changes in structure and content that are required in order for economic theory to come into line with the realities of the 21st century. Section 9 will propose a modest alternative starting point for modeling economic behavior, and section 10 will consider some of what this means for the teaching of economics.

Overall, the paper will focus on economic theory as it is taught in colleges and universities, because that is the main source of understanding of the economy for most of the population, including politicians and policy-makers.

1. The content of standard introductory economics courses

Economics textbooks are not only written for students. At two critical points in the history of economic thought textbooks have played significant roles in defining the field, not only for what is taught, but more importantly (in terms of real world outcomes) for the understanding of the economy that is used by politicians, policy makers, and the public, when it votes its approval or disapproval of how the government is affecting the economy.

This started in the 1890s, when Alfred Marshall wrote the first edition of his text, called *Principles of Economics*. It went through 8 editions, the last being published in 1920. For a large part of the English-speaking world Marshall’s textbook continued to define the field (especially the microeconomics basics) until the middle of the 20th century, when it was replaced by Paul Samuelson’s *Economics* (first published in 1948). That set the standard for about the next 60 years.

These textbooks have not only defined economics for students, they also set clear standards

for how people in general should think about the economy, having great influence on government policies and also on the economic research that supports policy. Samuelson was well aware of the impact of his texts, saying in an introduction, "I don't care who writes a nation's laws – or crafts its advanced treaties – if I can write its economics textbooks." Every year about 5 million people in the U.S. graduate from college having taken at least one economics course. These courses, and the textbooks that shape them, in turn contribute to a shared understanding of how things work in the world – and to a general consensus on whose voices will be heard on economic subjects.

Consider, for example, what might be called "Obama's dilemma": If you are not, yourself, deeply immersed in economics, how do you select economists to advise you on policy? Should President Obama have chosen his advisors based on reputation in the discipline? Or on personal economic success? I don't pretend to know in any detail what political constraints or motivations would have dictated what the president asked of his advisors, but it appears that he relied on both of these screens. Based on the first screen – reputation in the discipline – his choices would not have included anyone thinking about the new economy of the future, for the discipline has closed ranks very firmly around insiders, providing little opportunity for academic outsiders to become widely known. The second standard – being good at making money – moved him to select a number of his advisors from Wall Street, which has in recent years been among the most lucrative areas for amassing modern fortunes. If economics is about money, then, the reasoning goes, financiers must know a lot about it.

Without trying, here, to assess the effectiveness, or the goals, of Obama's economic policies, I will simply note that the problems in the economy that led to the 2007 crash and the Great Recession have not been solved; bubbles keep building up, enriching some people in the short run, and creating the potential for, once again, severe economic suffering for "main street" in the not so distant future. Neoclassical economic theory has failed to anticipate a number of severe problems that have been building up over the time of its intellectual dominance. To name just a few of the trends that have resulted from the system supported and celebrated by neoclassical economics, these include:

- ever greater income and wealth inequality;
- ever greater concentration of economic and political power in ever larger corporations – with severe negative impacts on the operation of democracy;
- a global climate that is rapidly changing in ways that threaten human health, the viability of many cities, the agriculture systems that feed humanity, and the diversity of plant and animal species on earth.

We are now at a time when economics is in need of another 60 year refresh. The heart of this need is in the question: How are human motivations and behavior to be understood in this human science? This paper will describe just one aspect of the new economic thinking that is laid out in the textbooks that I and my colleagues have written¹. It will work toward a model of

¹ Goodwin et al, *Microeconomics in Context, third edition; Macroeconomics in Context, second edition, and Principles of Economics in Context, first edition* – 2014, published by M.E. Sharpe. The other authors on the current editions are Jonathan Harris, Julie Nelson, Brian Roach, and (on the macro text) Mariano Torres. Earlier editions included Thomas Weisskopf and Frank Ackerman as my co-authors, with significant contributions by Kelvin Lancaster. The micro text has been translated into Italian, and, in a Transitional Economies edition, into Russian and Vietnamese.

economic behavior that considers how real people select goals; what we know about how behaviors are influenced by goals; and the limitations and influences that constrain both choice and action. I hope this can provide a good start to an alternative to utility theory, as a basis for understanding and presenting enough essential facts about human beings to support a useful and realistic economic theory.

Before considering an alternative I will briefly survey some of where we now are.

2. How we got here: Adam Smith minus Karl Marx; Keynes tortured by Samuelson

Adam Smith, generally regarded as the begetter of modern economic theory, stressed issues of growth and distribution, based on an image of smoothly functioning markets. The pieces of Smith's legacy that remained significant for what I will refer to as 20th century economics (though I will focus especially on the second half of the past century) were the emphasis on *growth*, and admiration for *markets*. This truncated legacy greatly reduced the emphasis on *distribution*, while also missing Smith's concern that markets might not always function optimally. He especially pointed to monopolistic behavior as a problem, and supported various kinds of government intervention to keep the market on track. Ignoring these caveats, 20th century economists pursued the optimistic program of modeling a world in which perfect markets lead to optimum social outcomes.

The classical economists – those holding the stage approximately until Marshall's time – also included Karl Marx, whose concerns for inequality and class conflict were shared by Smith (though they expressed themselves very differently). Marshall's deepest concern was with poverty, and the ways that (as he saw it) the poor were deprived of the means to develop their mental and moral capacities – what might be called, today, their human capital. Moral concerns were shunted aside in the positivism that overtook the field after Marshall's time. What remained of classical economic thought in the 20th century development of microeconomics was Adam Smith minus Karl Marx.

What of macroeconomics? We can see early strands in the work of the physiocrats (Smith went to France to study with them) who were concerned with issues of the balance of trade between nations – concerns that Ricardo took up a little later. Marx had a special focus on the macroeconomic instability of markets and he also, along with Ricardo, Malthus, and John Stuart Mill, raised some of what continue to be the critical macroeconomic questions:

- How is the total wealth generated by a society divided between those who own the means of production and those who work for them?
- Is the existing division optimal?
- What are the forces that determine how society's wealth will be divided?
- And: what are the goals of the economy?

None of these questions have been in the foreground of 20th century economics. Instead, in the second half of that century the field developed almost as though no one aside from

In addition to my debt to all the co-authors on the various editions of the "contextual" textbooks, I especially want to acknowledge Brian Roach for materials he contributed to Chapter 6 of *Microeconomics in Context, third edition*, which appear throughout this paper, especially in section 5.

Ricardo had ever thought about macro issues until Keynes came along, to prescribe how to get out of the Great Depression of the 1930s.

From the point of view of those who began to call themselves *neoclassical* economists there were some problems with Keynes' prescriptions. Politically, Keynes was on what came to be the losing side, in his conviction of the importance of government's role in stabilizing economies. Methodologically as well his approach did not fit the Procrustean bed on which Paul Samuelson laid the thinkers he worked over to come up with his *Principles* texts. The bits that hung over the bed and had to be chopped off included a belief in the probability of market failures. The bits that were too short and had to be stretched to fit Samuelson's passion for the tidiness and precision of mathematics were any ideas that could not easily be described in formal models.

Thus, if the skeleton of 20th century microeconomics was Adam Smith minus Karl Marx, that of macroeconomics was Keynes tortured by Samuelson.

3. Mainstream economics teaching in the late 20th Century

There are some true and useful things to be learned in standard 20th century economics, such as the basic concepts of supply and demand intersecting to create wages and prices. However if you ever took an economics course you may have since discovered that many other things also affect prices, such as advertising, or consumers' lack of information. And wages involve even more complicated human interactions, habits and expectations. These complexities and exceptions don't get much hearing in introductory courses – and, surprisingly, they get even less at the upper levels, where, instead, progressively more mathematics are imposed on a progressively more abstract picture of an economy. Meanwhile the students are also being taught a lot that is dangerous. Here are some of the take-aways from standard economics course:

- We don't need to worry about material resources – the price system and human ingenuity ensures that all resources are directed to their most valuable uses (with "value" determined by ability to pay).
- Concentration of economic power is not much of a problem. Its entanglement with political power doesn't merit any attention at all.
- Increased consumption (regardless of the content) is the primary measure of wellbeing.

About 40% of college students in the United States take at least one economics course. Students who, two years later, have forgotten the diagrams and equations, are likely to still retain an impression that only selfishness is rational, that limitless greed is a universal human characteristic, and that economic success – of a nation, or an individual – can be assessed strictly in terms of the dollar value of consumption. Beliefs like this are the background for a culture that will accept as perfectly normal Ponzi schemes and cooked accounts, tax fraud and tax havens, the exploitation of children, women and immigrants, and corporate expenditures to get the most favorable political environment. Institutions – from governments and legal or banking systems, to the institutions of the family or formal education – are

shaped by socio-cultural norms whose roots can be traced, in some significant part, to the standard teachings of economics – especially what it says about human psychology.

In recent decades the sorry state of the economic culture (e.g., on Wall Street) has repeatedly bubbled up into disasters. But the economic culture has not yet changed in significant ways – and neither has the economic theory beneath it. This poses a significant challenge. The economics profession is one of the most tightly closed in all of academia. Economists who write about different ways of understanding the economy don't get published in the mainstream journals. Faculty who disagree with the mainstream generally don't get tenure.

Fortunately, as the core of the profession has continued to marginalize those who disagree, there has come to be a growing outer circle of hyphenated economists: institutionalist-, Keynesian-, ecological-, feminist-, radical-, social, socio- economics, and many more. Some of the better known names associated with alternative views include John Kenneth Galbraith, E.F Schumacher, and Herman Daly, as well as Wassily Leontief, George Akerlof, Joseph Stiglitz, and Amartya Sen; the last four are among the Nobel prize winners who continue to pose serious challenges to the mainstream. The closed ranks of academic economists has been able – thus far – to keep out even those who have received such recognition in the world (creating, for example, Obama's dilemma). However the alternative voices are increasingly being heard – especially the new group of behavioral economists. This paper aims to provide a summary, and a little additional forward motion, for some of these critical alternative ideas.

4. The psychological “foundations” for neoclassical economics

When I was beginning my studies in this field economist Robert Solow commented to me that the great strength of economics is that it is *fully axiomatized*; the entire edifice can be deduced from the basic rationality axiom, which says that *rational economic man maximizes his utility*. The origin of this axiom is often traced back to Smith, whose most widely quoted phrase comes from a passage in which Smith approvingly notes that merchants take what, today, we would call, a protectionist position – doing so, not with any thought for the good of society, but because their security and profit is tied to domestic industry. Thus, he says, the merchant “is in this as in many other cases, led by an invisible hand to promote an end which is no part of his intention.”² Excerpts such as this have been used as a justification for the 20th century economic model's vision of an ideal world in which a society comprised of entirely self-interested economic actors would make the society as a whole better off, and the idea that

² To give the flavor of the full quote: "As every individual ... therefore, endeavours as much as he can, both to employ his capital in the support of domestic industry, and so to direct that industry that its produce maybe of the greatest value; every individual necessarily labours to render the annual revenue of the society as great as he can. He generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it. By preferring the support of domestic to that of foreign industry, he intends only his own security, and by directing that industry in such a manner as its produce may be of the greatest value, he intends only his own gain; and he is in this, [as in many other cases] led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest, he frequently promotes that of the society more effectually than when he really intends to promote it." Adam Smith, 1982, *The Glasgow edition of the Works and Correspondence of Adam Smith*, Oxford University Press. vol.2a, p. 456.

As an example of the widespread misuse of Smith's writing, it is interesting to note that people often refer to the “invisible hand” in arguments that cite Smith as a *proponent* of free trade – ignoring that Smith's use of the phrase speaks approvingly of protectionist merchants.

pursuit of self-interest is the *only* thing that is done by rational economic actors – and that anything else is *irrational*.

When Alfred Marshall set out to codify the ideas of the economists before him, his starting point regarding human nature was essentially the same as Smith's, with one interesting addition: Marshall took cognizance of a particular group of humans – economists; a group that did not exist, as such in Smith's time. Though Marshall did not say so directly, it is evident from his writings that he assumed that the motivation for this particular group was to improve the human condition; specifically, to reduce poverty so as to allow people to develop their higher moral and intellectual faculties, rather than being condemned to lives of desperate effort for simple survival.

Traces of this optimistic view of economists' motivations can be found in most texts since Marshall's time, but they were increasingly buried beneath the far more pessimistic and narrow view of human nature in general that came in with Samuelson (even though Samuelson, as a person, would fit reasonably well within Marshall's optimistic view about the character and motivations of economists).

The problem was the old desire, stemming from the beginning of the 20th century, to make economics truly a science, in the model of physics. As Philip Mirowski has spelled out³, neoclassical economics clung to a physics template from the 19th century which natural scientists had mostly discarded by the early 20th. Among the problems with that template (and they were many) was a positivist view of knowledge – a view that physicists themselves largely abandoned as they confronted the indeterminacies rife in quantum mechanics, general relativity, chaos theory, Heisenberg's uncertainty principle, etc.. The natural sciences have largely come to recognize that their practitioners are human beings, who have values, and that value-free science is virtually impossible. Neoclassical economics got stuck in an attachment to mathematics as the way to ignore the roles of values, history, institutions, politics, and other inconvenient subjects. Mathematics has much to offer to economics, but it is unlikely to find its best use when thus employed as a means of denial.

5. Behavioral economics

Neoclassical economics claims to be based entirely on a view of human nature which is not only morally repugnant, but which also both leaves out a great deal about how people actually do operate, while it brings in seriously contrary-to-fact assumptions about what people are capable of. The latter have included assumptions about consistency (including that preferences change slowly, if at all, and that if A is preferred to B and B is preferred to C, then C cannot be preferred to A); about information (people are able to act as *if* they have perfect information); about self-knowledge (people know what they want, and are best served by getting what they want); and about influence, or power. The last of these assumptions includes the idea that human wants and preferences are endogenous, generated entirely from within; it ignores the extent to which people's choices and decisions may be manipulated by those who have an interest in persuading the public to buy certain things, or vote in certain ways. It ignores the reality that market economies are rife with powerful actors who do have such an

³ In *More Heat Than Light: Economics as Social Physics, Physics as Nature's Economics*; Cambridge University Press, 1989

interest, in both the economic and the political spheres.

A paper on the sociology of economics⁴ would describe how these unrealistic assumptions have been fostered in a profession with skewed motivations. Promotion and tenure in college and university economics departments depend on publication in a short list of acceptable journals. The editorial boards of those journals have an interest in keeping the ideology unchanged, for several reasons: Their status depends in part on the mystification of arcane language and hard-to-swallow assumptions; these characteristics, as well as emphasis on difficult mathematics, erect barriers to entry to the profession; control over the supply of economists results in an ability to command higher salaries than most other academics, as well as the possibility of much higher pay in the service of business or politics, where there is also an interest in maintaining the status quo⁵.

In the last few decades the narrow economic view of human behavior has been challenged by a strong alternative called behavioral economics. Studies in this area suggest that a more sophisticated model of human motivations is required to explain such behaviors as those that lead to stock market swings, the ways that people react to good and bad fortune, and why people often seem to act against their own self-interest.

Perhaps the most famous contemporary behavioral economist is not an economist at all. Despite being trained as a psychologist, Daniel Kahneman, along with his frequent colleague Amos Tversky, won the 2002 Nobel Memorial Prize in economics. Kahneman's research has found that people tend to give undue weight to information that is easily available and vivid – a detour from 20th century assumptions of economic rationality that he calls the “availability heuristic.” For example, suppose college students are deciding which courses to take next semester, and they see a summary of evaluations from hundreds of other students indicating that a certain course is very good. Then suppose they watch a video interview of just one student providing a negative review of the course. Even when students are told in advance that the negative review was atypical, they tend to be more influenced by the vivid review than the summary of hundreds of evaluations.

Kahneman has also shown that the way a decision is presented to people can significantly influence their choices, an effect he refers to as “framing.” For example, consider a gas station that advertises a special 5-cent per gallon discount for paying cash. Meanwhile, another station with the same prices indicates that they charge a 5-cent per gallon surcharge to customers paying by credit card. While the prices are exactly the same, experiments show that consumers respond more favorably to the station advertising the apparent discount.

An effect similar to framing is known as “anchoring,” in which people rely on some not necessarily relevant piece of information as a reference point in making a decision. In a real-world example, a high-end kitchen equipment catalog was selling a particular bread maker for \$279. Sometime later, the company began offering a “deluxe” model for \$429. While they did not sell many of the deluxe model, sales of the \$279 model almost doubled because now it

⁴ See, e.g., Goodwin, Neva, 2008 “From Outer Circle to Center Stage: The maturation of heterodox economics”.

⁵ To give just one example of the latter point: standard utility theory, as portrayed in 20th century economics, *could* be used to show that a high degree of economic inequality is bad for economic stability, and reduces overall well-being. This conclusion is too rarely drawn in the standard literature, even though a few writers such as Robert Frank make plain the logic.

seemed like a relative bargain.

A conventional view of rationality is that emotions get in the way of good decision making, as they tend to interfere with logical reasoning. Again, however, research from behavioral economics suggests a more nuanced reality. Logical or rational reasoning is most effective when making relatively simple economic decisions, but for more complex decisions we can become overwhelmed with too much information. For example, Ap Dijksterhuis, a psychologist in the Netherlands, surveyed shoppers about their purchases as they were leaving stores, asking them how much they had thought about items prior to buying them. A few weeks later, he asked these same consumers how satisfied they were with their purchases. For relatively simple products, like small kitchen tools or clothing accessories, those who thought more about their purchases tended to be more satisfied. But for complex products, such as furniture, those people who deliberated the most tended to be *less* satisfied with their purchases.

6. Bounded rationality

Rationality has become a loaded word in economics, bringing with it the baggage of earlier models that did not anticipate the findings of behavioral economics or take into account other every-day observations. The traditional rationality model includes the assumption that rational behavior is optimizing behavior (“rational economic man *maximizes* his utility”). In the 1970s an extreme version of this made the further assumption that rational economic actors have “perfect information.” A slightly more modest version says that people will collect information until the perceived costs of acquiring additional information exceed the perceived benefits.

One of the most effective challenges to the traditional assumption of rationality came from Herbert Simon, another non-economist winner of the Nobel Memorial Prize in economics (1978). Considering whether it is indeed possible for people to identify the optimal point at which one should cease gathering additional information, Simon showed that one first needs to have complete knowledge of all possible choices in order to identify that optimal point. Determining what additional information might be out there, and then gathering it, can be very costly in time, effort, and money – if it is even possible. Accordingly, Simon maintained, people rarely optimize. Instead they do what he called “satisficing;” they choose an outcome that would be satisfactory, and then seek an option that at least reaches that standard.

Given constraints of time and other resource limits, satisficing seems to be a reasonable behavior. If an individual finds that the “satisfactory” level was set too low, a search for options that meet that level will result in a solution more quickly than expected, or perhaps even multiple solutions. In this case, the level may then be adjusted to a higher standard. Conversely, if the level is set too high, a long search will yield nothing, and the “satisficer” may lower his or her expectations for the outcome.

Another deviation from rational behavior as traditionally defined has been called “meliorating;” this may be described as starting from the present level of well-being and then taking any opportunity to do better. A simple example is a line fisherman who has found a whole school of haddock but only wants to keep one for his supper. When he catches the second fish he compares it to the first one, keeps the larger, and throws the other back. Each subsequent catch is compared to the one being held in the bottom of the boat. At the end of the day, the

fish he takes home will be the largest of all those caught (and the sea-gulls will have become very fat!).

One result of using melioration as the real-world substitute for theoretical optimization is its implication that *history matters*: People view each successive choice in relation to their previous experience. It is commonly observed, for example, that people are reluctant to accept a situation they perceive as inferior to previous situations. This psychological path dependence – the way feelings about the future depend on previous experience – is relevant to how people feel about rising prices, and even more so to attitudes toward declining wages.

Satisficing and meliorating may both be included under the term “bounded rationality.” The general idea is that, instead of considering all possible options, people limit their attention to some more-or-less arbitrarily defined subset of the universe of possibilities. Usually these subsets consist of the options immediately evident, along with others specifically sought out through some simple decision rule. For example, when deciding what to spend her money on, an individual may at one time confine her consideration to “major expenditures,” such as a college education or an apartment; at another time she might contemplate “expenditures on food”; and at another time she might sit down to work out budget categories, pondering, for example, “How much should I spend on food each month, how much should I devote to entertainment, and how much shall I set aside for a major need like an apartment?” With satisficing or meliorating behavior, people may not choose the “best” options available to them, but they at least make decisions that move them toward their goals.

7. The role of influence

Herbert Simon received the Nobel Prize in 1978. This fact had little or no influence on subsequent economics textbooks, which sometimes mentioned bounded rationality, but did not reduce their dependence on the old rationality postulate as the foundation for deducing all human behavior.

Simon was not the first critic to be so dismissed. Decades before behavioral economics came into fashion “alternative” economists were complaining about the unrealism of the neoclassical view of humanity. They especially focused on the fact that, as Smith had so well recognized, people are social animals. Relatively few of our actions are taken completely without regard for what we have seen other people do, or what we expect that other people will think. Even popular books on finance refer to the “herd instinct” in reference to the way investors follow fads and fashions of thought. There appears to be an inborn tendency for people to act as part of some kind of human collective, rather than in isolation. Yet this had no place in the neoclassical understanding of human behavior.

Since the social nature of human beings has been discussed at great length in writings from many disciplines, the remainder of this section will focus on a different aspect of our social embeddedness that has been seriously overlooked in neoclassical writings: The intentional influence exerted by people with a political or a sales agenda.

Behavioral economics has shed light on a number of ways that others can affect our decisions, by setting a “frame,” or providing extra emphasis on one conclusion at the expense of others. Available information is, of course, a critical feature, and actors other than the

decision maker may have a strong influence on what information is available. These, and other ways that decisions may be distorted by influences not related to the goals of the particular actor, have long been well known to politicians and advertisers, who, since the early part of the 20th century, have often based their successes on assuming *irrational* consumers and voters.

For example, food companies are well known to cater to people's innate physical attraction to sugar, fat and salt. These three elements support health when eaten in appropriate amounts, but they were rarely available in sufficient quantity during most of human evolution. We are all therefore born with some degree of craving for these substances, yet our health suffers if we do not recognize when we have had enough. Television teaches children at a very early age to recognize logos that are associated with sugar, fat and salt. Kids then often become loud advocates in grocery stores for the products that will keep them from learning how to recognize "enough." Obesity and other forms of ill health are the obvious results. These outcomes could be mitigated if parents and other consumers were acting on a rational, well-informed understanding of how health is affected by food choices. That they often do not do so may result from a combination of poor (frequently misleading) information, persuasive marketing, and short-term desires (to answer a sudden craving, or quiet a screaming child) that override long-term understanding of cause and effect.

On the side of politics, we increasingly witness voters supporting political platforms that are against their own interest, or simply impossible. To give just a couple of examples in the U.S. context: The states that receive substantially more federal monies than they pay in federal taxes are, on the whole, the ones whose voters are most anti-government. Kansas is the state that (because of frequent tornadoes) is the third largest beneficiary of aid from FEMA (the Federal Emergency Management Agency), yet Kansans vote for representatives who vote to reduce funding for FEMA. Gail Collins, op-ed writer for the New York Times, notes that the current governor's race is dominated by denunciations of federal spending on government activity – even while Alaska gets more federal money per person than any other state. What's especially striking about such public discussions is their denial of reality. As Collins continues, "there's virtually no discussion of eliminating anything its residents – who pay no state income tax or sales tax – get now," including "Alaska's super-subsidized mail service." (August 16, 2014). This position, being popularized by potential state governors, is not only against the interests of the people, it spreads a false belief that you can get what you want from government while tearing it down.

A central belief of the ideology of 20th century economics has been that individuals are always the best judge of their own well-being, and of what will contribute to it. This has made it impossible to incorporate anywhere within the theory a recognition of how economic and/or political power can influence individual goals, choices and actions. What would it mean for economic theory to recognize this reality? What use would it make use of evidence that institutions and policies can be constructed to encourage people to make better decisions?

A 2009 book titled *Nudge: Improving Decisions About Health, Wealth, and Happiness*, written by an economist and a legal scholar, suggests that governments and other institutions could and should play a role in promoting better decision making. For example, the authors use a cafeteria as an example of a setting in which people might be encouraged to make "better" decisions if, say, healthier options at the salad bar were placed at easier reach than the less healthy alternatives. A growing recognition that corporations are promoting bad food choices

has added to the effects of the 2007-8 recession to motivate some rethinking of the “markets-are-always-the-best-place-for-individuals-to-make-decisions” ideology. An alternative view that is beginning to get a hearing is another recent New York Times op-ed in which Mark Bitman says:

If the most profitable scenario means that most food choices are essentially toxic – in the sense that over-consumption will cause illness – that’s a failure of the market, not of individual choice. And government’s rightful role is not to form partnerships with industry so that the latter can voluntarily “solve” the problem, but to oversee and regulate industry. Its mandate is to protect public health, and one good step toward fulfilling that right now would be to regulate the marketing of junk to children (June 18, 2014).

8. Ethics, goals, and well-being

Twentieth century economics supported, implicitly when not explicitly, the idea that neither ethics nor history nor the institutions of law or culture were of much economic importance – as long as these things did not get in the way of “free” market functioning. This case was pressed with special vigor from about 1970 to the end of the 20th century by economists from what was known as the Chicago School.

Even early on in this period there began to be concern that individuals acting solely to achieve their personal goals could not be counted on to operate a business in ways that would be good for the business itself. This real-world concern, combined with the dogma that people only act on the basis of self-interest, resulted in various efforts to motivate business leaders by offering rewards for specific markers of success (such as the price of the company’s stock). These efforts had the unintended consequence of escalating compensation of top management in the United States to levels that were many times greater than anything that had previously been considered normal (or were normal in other countries). They also resulted in an increasingly short-term vision on the part of business leaders. Very large scale frauds, Ponzi schemes, tax evasions, and environmental and human costs that businesses externalized during this period have made it increasingly evident that society cannot afford to encourage a culture of economic activity that ignores all normal human motivations except the selfish pursuit of personal gain.

With the advent of behavioral economics, and the various streams of psychology that have fed into it, there is increasing recognition for an alternative position, that a well-functioning economy cannot rely only on self-interest. The notion of “social capital”,⁶ which began to gain traction in the 1990s, formalized the idea that, without ethical values that promote trust, inefficiencies would overwhelm any economic system.

Absent such values as honesty, for example, even the simplest transaction would require

⁶ See Goodwin, “Five Kinds of Capital”; also "[The Limitations of Markets: Background Essay](#)." It is worth noting that one of the most famous institutionalists, Gunnar Myrdal, was co-awarded a Nobel Prize in 1973 for explaining why values are always with us. The current lack of attention to institutionalists, including those who have received such attention, is another example of the ability of neoclassical economics to marginalize ideas, and their proponents, that do not fit within the rigid neoclassical paradigm.

elaborate safeguards or policing. Imagine if you were afraid to put down your money before having in your hands the merchandise you wished to purchase – and the merchant was afraid that as soon as you had what you wanted you would run out of the store without paying. Such a situation would require police in every store – but what if the police themselves operated with no ethic of honesty? If everyone in business cheated whenever they thought they could get away with it, business would grind to a halt. If everyone in the government worked only for bribes, meaningful governance would disappear. And it is hard to imagine how the human race would survive if altruism was not common enough so that people would be willing to make sacrifices of time, convenience and resources to meet the needs of those who cannot take care of themselves, such as children or sick people.

Among economists some attention is again being paid to the fact that many real-world problems would be difficult, if not impossible, to solve if there were not in fact a reasonable number of people willing to work for the common good – the general good of society, of which one's own interests are only a part. Fortunately, recent experiments on human behavior demonstrate what most people who are not blinded by models of "rational economic man" have realized all along: That people really *do* pay attention to social norms, and they are willing to reward those who follow these norms and to punish people who violate them, even

when this has a cost in terms of their narrow self-interest.⁷ This point has great importance for a discipline that has the potential for affecting social norms. People who have studied economics in recent decades have carried away from those studies, into the wider culture, messages that only selfishness is rational, altruists are suckers, and one does not need to think about goals or values to know that private enterprise is always more efficient than – and therefore preferable to – any kind of collective action, including government.⁸

Economics, over the last 60 years, has set itself directly at odds with the basic ethical concerns of all major philosophical and religious teachings. In this respect economics is an inferior guide. From the point of view of society as a whole, purely selfish behavior will often

⁷ A well-known example from behavioral economics is the "Ultimatum Game" in which two people are told that they will be given a sum of money to share, say \$20. One player gets to propose a way of splitting the sum. This person may offer to share \$10 with the second person, or only \$8 or \$1, and plan to keep the rest. The second person cannot give any input to this decision but can only decide whether to accept the offer or reject it. If the second person rejects the offer, both people will walk away empty-handed. If the offer is accepted, they split the money as the first person indicated. If the two individuals act only from narrow financial self-interest, then the first person should offer the second person the smallest possible amount – say \$1 – in order to keep the most for him or herself. The second person should accept this offer because, from the point of view of pure financial self-interest, \$1 is better than nothing. In fact, however, researchers have found that deals that vary too far from a 50/50 split tend to be rejected. People would rather walk away with nothing than be treated in a way that they perceive to be unfair. In the context of social relations, even the most selfish person will gain by serving the common good and thus walking away with somewhere around \$10, rather than just looking at his or her own potential personal gain and quite possibly ending up with nothing.

⁸ A number of studies have shown that economics students and faculty are less altruistic than others. In one example, economics students expressed a lower willingness to contribute money to pay for public goods than other students. The same was found of economics faculty, in spite of their average pay being higher than the faculty in the other disciplines to which they were compared. (Bauman, Yoram, and Elaina Rose, 2011.) Similarly, "...researchers who undertook a number of free rider/prisoner's dilemma games, found students with a training in economics to be more aggressive, less cooperative, more pessimistic about the prospects of cooperation, and more prone to cheating than students who had not undertaken any economics subjects (note that selection bias was controlled for in these experiments). The characteristics that developed as a result of taking these economics courses persisted long after their education had finished." (Frank, Gilovich & Regan 1993, 1996, cited in Thornton, 2013.

fail to promote social well-being. Economists are finally beginning to recognize this reality, first with the reluctant admission that externalities do exist, such that market outcomes (often equated with the invisible hand) do not reflect all the impacts of market behavior, as they would do in the ideal, perfectly functioning market. Even the economic actors themselves – whether they are business people, individuals acting in their family or community roles, or governments – may lack the information needed to make what 20th century economics assumed as the rational decisions that would lead to social optima.

Economic theory, and the textbooks through which the theory is summarized and passed on, need to catch up to these realizations. A good start would be to broaden the debate on goals. In the 21st century it is increasingly evident that ecological problems and constraints are coming into serious conflict with the goal of maximizing GDP, for any one country, and especially for the world as a whole. A more appropriate goal for our time could be stated as: *To maintain and increase human well-being, without further harm to the ecosystem.* (The final clause of that goal statement could be rephrased as ...*without increasing consumption of the high-end goods now typical in rich countries.*) This may be followed with a further proposition: *An important goal of the discipline of economics should be to help people understand how to move their economy toward its goals.*

If or when such a shift in goals occurs it will dramatically alter a good deal of what is taught in economic textbooks. Among other things, if the well-being that we would aim to support cannot be defined concretely and quantitatively enough to lend itself to the use of the calculus, can or should we be talking about *maximizing* well-being? Or is a subtler approach required – one that does not posit objectives that can be weighted into a single maximand, but that is prepared to use judgment to deal with tradeoffs? (The issue of judgment will be discussed in the last section of this paper.) Other questions raised by the adoption of more complex goals include: What kind of economic growth or development can promote present well-being while preserving productive resources for the future? Can we imagine changes in values and in the economic culture, as well as the broader culture, that will make it easier to promote the most well-being-serving growth or development? How are the answers to these questions different for rich vs. poor countries?

These difficult questions are not discussed, but are glossed over by an implicit assumption discernable in 20th century economics texts: That an economist has, and can turn to, a client – whether this is an individual or a maker of national policy – who has a clear idea of his or her goals. In real life, outside of textbooks, *macroeconomists* do frequently have clients, who present them with questions into which goals may be read – but often the client (such as President Obama in 2008) is hoping that the economists will help to clarify the goals and the priorities. If the overriding goal is “Get the country out of this mess!” should the first priority be to save the banking system, or to protect jobs, or to keep people from losing their homes? Is there a necessary order in which these problems must be tackled? Obama’s team came up with one set of answers and priorities; a different group of economists would have defined the question, the goals, and the priorities, differently.

This means that economists are not off the hook. Their values, and the goals that arise from them, are inevitably relevant, not only for the advice they give to heads of state, but also for many smaller tasks – and, importantly, for how they teach economics in schools and institutions of higher education. Unless they have a client whose goals are unusually well-defined, macroeconomists still need to ask, Who speaks for society? When democracy is

working well, there are discernable answers to that question; when it is not, the economist will more often be left to define a large part of the question, as well as the answers.

There is a tradition in microeconomics of assuming that individuals are the best judges of what will provide them with well-being, with the exception of young children and the mentally ill, who often fail one test of rational goal selection: that is, to select goals such that, when they achieve them, they will be glad in the long run that they have done so. Overall, even while assuming that more consumption is always more desired than less, economists have been wary of commenting on the goals people set. Yet recent research has indicated that the happiness people experience in life is strongly related to the goals they set. This is relevant to economics if happiness, as a component of well-being, is a goal for an economy. Anthropologist Tim Kasser, economist Robert Frank, and others working in the area of hedonic psychology show happiness and mental health to be negatively correlated with strongly materialistic goals, especially when the goals are set in relation to others' achievement (i.e. the goal is to have something more or better than ones reference group). This idea is not new. Alfred Marshall assumed that the moral structure which is part of the foundation for individual motivations is, or should be, one of society's most important ends: The ultimate public good lies in a kind of progress wherein human wants are educated so that individuals will increasingly want what is good for them. What is good for people, Marshall felt, is to want the kind of reward that a good person wants: i.e, distinction, honor, and the pleasure, for its own sake, of serving others. If the moral structure of society and of its individual participants can gradually be brought to this orientation the whole society will be better off, for honor could partially replace pay as the reward at the higher levels of work effort, permitting an even distribution of income without loss of productivity; and consumers as well as workers will be better off, as individuals at every level take more pride in the quality of their work.⁹

Tibor Scitovsky, in *The Joyless Economy*, contrasted Americans' pursuit of pleasures that do not require effort to Europeans who, as he saw them (from a mid-20th century perspective) expected to put in effort to learn to enjoy, for example, challenging works of art, whether in music, writing, or other forms. Amartya Sen attempted to formalize this notion in his concept of a "two stage utility function" wherein he imagined that first people decide what kinds of utility are involved in a given problem (i.e., are we after the utility we will feel by doing our duty; by that associated with self-improvement; or is it simply hedonistic pleasure?) Having made this choice, we then choose the activity that will maximize the preferred type of utility.¹⁰ Albert Hirshman cogently remarked that:

Men and women have the ability to step back from their "revealed" wants, volitions and preferences, to ask themselves whether they really want these wants and prefer these preferences, and consequently to form metapreferences that may differ from their preferences....

When a change in preferences has been preceded by the formation of a metapreference... it typically represents a *change in values* rather than a change in tastes. ("Against Parsimony: Three Easy Ways of Complicating some Categories of Economic Discourse" 1984. Italics in the original.)

⁹ Alfred Marshall, 1907, "The social possibilities of economic chivalry"

¹⁰ Amartya Sen, 1977, "Rational Fools"

Behavioral economics derives its view of human nature from observations of behavior, often under carefully controlled experimental conditions. The neoclassical view derives all expectations of human behavior deductively from the rationality assumption. For a period in the 20th century it seemed that evolutionary theory gave scientific support for the latter approach, when early writings in sociobiology suggested that the individual survival imperative would always prevail over any other motives. In the latter decades of the century this simplistic view was strongly rebutted by other sociobiologists who pointed out that even the most “selfish gene” operates so as to promote the future continuance of the group that carries this gene. This may be seen in action, for example, when birds court danger as they try to lure a predator away from their young. But other-regarding behavior goes beyond simple gene preservation, as in the many stories of human heroism which illustrate human choices to sacrifice individual survival for the sake of other people, whether or not they are genetically related.

9. The model of economic behavior in contextual economics

This paper has described how neoclassical economics has managed to stretch, shrink, or ignore ideas, such as those of Karl Marx, Adam Smith, and John Maynard Keynes, as well as the institutionalists, femanists, ecological economists, etc., wherever these ideas threaten the essence of the neoclassical paradigm. Can we hope that behavioral economics will finally be able to drive home the points that Simon, Kahneman, Sen, Hirschman, Myrdal, etc. were making over much of the last century? A somewhat discouraging view on this is given by Tim Thornton in his impressive doctoral thesis on possibilities for changing economics curricula:

Earl (2010), in part drawing on the work of Sent (2004), makes the point that what now passes for behavioural economics — what he terms the ‘new’ behavioural economics — is in key ways a betrayal of the ‘old’ behavioural economics founded by Simon (1957). The existence of a ‘new’ and ‘old’ behavioural economics is actually part of an established pattern, where we have a ‘new’ and ‘old’ institutionalism (Rutherford 1994) and Keynesian and post-Keynesian economics (King 2002). The paradigmatic strictures of orthodoxy prevent a proper engagement with, and understanding of, the original ideas (Earl 2013). This means that new ideas are only absorbed into the mainstream on the mainstream’s own terms, leaving most of the original and more challenging ideas to exist only within marginalised and largely ignored schools of economics.

The conservative nature of much of the ‘new’ behavioural economics is evident enough in the assertions of its key contributors and textbooks. Behavioural economics is seen as an approach that “extends rational choice and equilibrium models; it does not advocate abandoning these models entirely” (Ho, Lim and Camerer cited in Wilkinson 2008 p.4). The prominent new behavioural economist Matthew Rabin is “adamant that he wants to create a sense of continuity that allows people to see the changes that are happening as incremental changes to a fundamentally unchanged science” (Bateman 2007 p.6). In summary, while the mainstream research frontier is different enough to confound conventional notions of the orthodox-heterodox

dualism, [it] is not a revolutionary force that is changing the face of economics.”¹¹

Thornton is certainly correct that behavioral economics is encountering the expected resistance, in the form of an apparent embrace – what might be termed a “smothering embrace” – which struggles to preserve the core of the old paradigm, even when that requires damaging the new ideas. Complexity economics (a subset of the various kinds of complexity theories that have excited and energized several branches of the natural sciences and mathematics in recent decades) has similarly been touted as an unstoppable force for change in the discipline, while also meeting subtle resistance to its potential for change. The same can be said for events in the world, outside of theory: the Great Recession (and the failure of economics to identify, before the fact, the economic trends that were leading to it), and the related recent emphasis on inequality. Might all of these forces together break through the barriers to real change in economic theory, teaching and practice? I cannot predict this, but am somewhat optimistic. And, in my optimism, I have written this paper to propose an alternative to the starting point of neoclassical economics.

I have suggested that the axiom “rational economic man maximizes his utility” does not stand up to tests of logic, evidence, or the needs of society. Twentieth century neoclassical methodology depended heavily on this radically simplified statement to underpin its boast of being scientific, in being “fully axiomatized.” Even mathematics no longer accepts that as a realistic requirement. In an economics designed to grapple with the complexities of the world in the 21st century it is neither necessary nor possible to formulate a statement that will be used as an axiom from which all the rest of economic knowledge and understanding can be deduced.

With that said, I propose the following statements concerning motivations and behavior as reasonable starting points for this human science.

Normal economic behavior must be understood within:

1. The social context: People care what others think and do. Individual actions are not only motivated from within, but also by a sense of group identity.
2. The ethical context: Experience and observation suggest that most people pursue a variety of goals, normally including some mixture of self-interest and concern for others or for the common good. Goals are not identical to revealed preferences; people’s values come into play in the actions they undertake and the goals they set for themselves, for their children, and for their society.

A reasonable definition of rational behavior includes:

3. Choosing goals such that (a) when the actor achieves the goals, she or he will be glad to have done so; and/or (b) the pursuit of the goal itself contributes to well-being.
4. Pursuing those goals in a manner that the actor reasonably expects will lead toward their achievement.
5. Limits to rationality: Most adults who are not suffering severe psychological or cognitive handicaps attempt to act rationally, as just described. However, sometimes

¹¹ Thornton, 2013. In a personal communication (July 2014) Thornton notes that he is generally hopeful about the prospects for a reformed economics, but argues that it will require the adoption of a diverse suite of strategies, as well as some adjustments in our understanding of the nature of the problem to be solved.

lack of information, or the influence of conflicting emotions, or influence from others pursuing different goals, may cause generally rational actors to choose goals that are not consistent with well-being, or to do things that lead away from their own goals.

Can this description of rational behavior, set in social and ethical contexts, and clearly looser than the old utility-maximizing axiom, be used as the basis for models that will look like the models held up by neoclassical economists as the ideal way to present the world for economic analysis? The following seems to me a good answer to this question:

... a more radical reformulation would discard altogether the idea that a universally applicable model, in which all key relationships are predetermined, can describe the economic world. Economic behaviour is influenced by technologies and cultures, which evolve in ways that are certainly not random but which cannot be described fully, or perhaps at all, by the variables and equations with which economists are familiar. Models, when employed, would therefore be context specific (John Kay, 2013).

What would economic theory, teaching and application lose, and what would it gain, by replacing the 20th century model of human psychology with the looser set of statements suggested above? The loss would be a good deal of the edifice of neoclassical economics – that portion that is in fact built upon the very narrow definition of human behavior, in economic contexts, as purely motivated by self-interest. This narrow definition had been encrusted with some further assumptions about rationality, including consistent preferences as well as the possession of information that was either “perfect” or “sufficient” to make optimally self-interested decisions. These assumptions have been shown to be invalid and must also be abandoned.¹²

A deep refresh for economic thought which begins by abandoning these assumptions opens up important new possibilities in several areas of content as well as methods, including many that are beyond this paper. I will conclude with comments on just one of the topics to which there may be a new opening: Consideration of economists as human beings.

10. The education of economists

This paper has been largely about the people whose actions are the subject of the discipline of economics. This final section will consider some issues to do with economists themselves. There is much to be gained by respecting each individual as a source of knowledge about human beings, and assuming that most individual economists possess (whether or not they draw on it) more knowledge about human motivations and behaviors than can be summarized in any “model” or other set of statements.

¹² Milton Friedman famously stated that it did not matter if assumptions reflect reality, as long as the theory based on them can make good predictions. Economics has dramatically failed to predict many important aspects of the last 50 years – from the unintended consequences of motivational pay at the top corporate level (based on principle-agent theory) to the repeatedly swelling and bursting bubbles in the 1990s and into the next decade. The assumptions of rational maximization have failed by Friedman’s test as well as on the reality tests imposed by behavioral economics. See Goodwin, 1991 “Stories That Blow up: How to Anticipate When the Realisticness of Assumptions Will Matter,” chapter 11 in *Social Economics: An Alternative Theory* (New York, St. Martin’s Press.)

The previous section proposed a basic summary statement of what economics, and economists, should know about people. Those who accept this summary as reasonable will also note that it is very far from being a full description of “what people are like.” Volumes and libraries can be devoted to that question without laying it all out. The only complete model of a human brain is that brain; computers full of algorithms cannot provide the information on you that is contained in your memory, intuition and knowledge.

Economists, like other people, are not simple calculating machines. Students entering the field bring to it their life experience and their personal qualities such as judgment, caring, hope, imagination, and intelligence (of many kinds). A good educational experience will build on these qualities – it will provide more and deeper knowledge and experience, ways of critically analyzing information, and ways of researching new information, both alone and working with others. These goals for education are quite well known.

While schools of education are concerned with the goals of education, economists have accepted very narrow goals – “learn what you are taught,” but not “think about the meaning and the application of what you learn.” I referred earlier to the convenient assumption that economists out in the world will have clients who will be able to define any necessary goals. Since the reality is often not so simple, perhaps writers and readers of macroeconomics textbooks should take on themselves the role of the client; for example the first and the last class meeting in any macroeconomics course could include a discussion of the appropriate goals for the economy.

Another, more specific goal that has not received sufficient attention within economics (though it turns up in some other places in higher education) is that of training judgment. This quality is, in fact, as essential for neoclassical economics, with its strong emphasis on techniques of analysis, as it is for a less formal approach. There is nothing inherent in any technique *per se* that assures that it will be used in the right place. The correctness of the conclusions, however they are drawn, will depend upon judgment at several points:

- Judgment is required, to start with, in deciding what topics are more or less interesting to analyze.
- Where any empirical facts are used, judgment is critically required for the selection of which facts to focus on.
- Different techniques applied to what appear to be the same materials are likely to produce different results – aside from the fact that the choice of technique may impose restrictions on the selection of the variables under consideration. (For example, hard-to-quantify variables may be ignored in a model that requires quantitative representation of its subject matter). It is a matter of judgment which is the most appropriate technique for the analysis of a particular subject.
- Any abstraction from reality, whether it is a formal model or a narrative chain of reasoning, requires some assumptions; judgment is required in their selection.¹³
- It is also required for the reading, evaluation, interpretation and application of the conclusions that emerge from any analysis.

¹³ Just as one example, when analyzing the likely effects of a tax an economist may draw on existing estimates of the price sensitivity of consumers. Judgment may be required to choose which existing estimate to use, or whether any estimate, often drawn from different situations, is good enough to use, or if it should be adjusted before being used.

Judgment can, and should, be informed by the gathering of information – but, as Herbert Simon pointed out, it is generally impossible to know exactly what is the information we need that we don't have, or when we have enough. Judgment also can and often should be assisted by analysis – but, as was noted above, complex decisions may actually be hindered by excessive ratiocination, and may be more effective when taken quickly, calling on emotion or intuition. Thus, to add to the list of bulleted points above,

- Judgment is required to know when to stop gathering information, and when to call on intuition as an aid to analysis¹⁴.

Humility is another valuable quality, too seldom found in economists. No individual can be expert in all the fields that are relevant to the important subjects for economics in the 21st century – fields such as ecology, systems theory, sociology, psychology, history, nutrition, anthropology, philosophy, political theory, etc. Each economist should have studied enough outside of economics so as to be able to talk comfortably with people from other disciplines, and to know where to look for insights that economics lacks. This means that economists have to talk in terms that non-economists can understand – i.e., not jargon.

Humility tells us that we don't know all the answers: imagination is required to find solutions that are not obvious. Imagination is the quality that may be hardest of all to teach, but at least economists could learn to recognize and value it.

Thus I come to a list of the qualities that would, ideally, be promoted in people who are learning to be economists. The list includes judgment, humility and imagination, as well as good communication skills – the ability to learn from others, and the ability to communicate with people with different educational and cultural backgrounds. Because humanity now confronts so many difficult problems, economists also need to be attuned to recognizing and defining problems. And along with humility they need to have hope and optimism that they can find solutions.

Finally there is the quality that existed in Adam Smith and Alfred Marshall – and, indeed, in Paul Samuelson – and in most of the other great economists: That is the quality of wishing to contribute to making the world a better place. Many students come to economics because they have this wish, but those who are most concerned with it are the ones who, in recent decades, have been turned off and dropped out of the field, disgusted with the lack of overt values of caring and community, and the presence of implicit values of greed and selfishness.

Neither I nor anyone else should insist on what “better” (as in “making the world a better place”) means to each person who studies economics – only that they should, ideally, be open-minded to the possibility that their notions of better and worse states of the world might change in the course of their education. Unlike imagination, the quality of caring is relatively

¹⁴ This last point may seem especially surprising, so it is worth giving an example. Recall the conclusion, earlier, that while ratiocination assists in good decision making in relatively small matters, it may be better to give room for intuition in larger ones. Thus, suppose an individual is presented with the decision about what research area she should choose for the next several years – or even for her life's work. Most economists who have both been successful and have enjoyed their life's work would probably agree that this decision was not made purely on a cold-blooded, rational basis. Once it is made, then intuition will often (but not always) take a back-seat to logical analysis.

easy to teach – by example; for it is intrinsic in people to pick up on and internalize what matters to influential people around us. Thus it is all the more important that those who teach economics investigate within themselves what matters to them, and how they think the economy, and economic theory and teaching, could contribute to a better world.

References

- Yoram Bauman and Elaina Rose. "Selection or Indoctrination: Why Do Economics Students Donate Less than the Rest?" August 2011, *Journal of Economic Behavior and Organization*, vol 79(3):318-327.
- Richard Dawkins, 1976, *The Selfish Gene* (Oxford University Press)
- Ap Dijksterhuis, Maarten W. Bos, Loran F. Nordgren, and Rick B. van Baaren, 2006, "On Making the Right Choice: The Deliberation-Without-Attention Effect," *Science*, vol. 311(5763):1005-1007 (February 17)
- Marianne A. Ferber and Julie A. Nelson, eds., 1993, *Beyond Economic Man: Feminist Theory and Economics* (University of Chicago Press)
- Robert Frank, 2010 (Paperback), *Luxury Fever* (Princeton University Press)
- Frank, Gilovitch, and Regan, "Does Studying Economics Inhibit Cooperation"
http://www.gnu.org/philosophy/economics_frank/
- Milton Friedman, 1953, *Essays on Positive Economics* (University of Chicago Press)
- Neva Goodwin, 1991 *Social Economics: An Alternative Theory* (New York, St. Martin's Press.)
- Neva Goodwin, 2005, "Five Kinds of Capital" at
http://www.ase.tufts.edu/gdae/publications/working_papers/03-07sustainabledevelopment.PDF.
Also "[The Limitations of Markets: Background Essay](#)," published by [CasePlace.Org](#) as part of a collection on [Market Failures: Corporate Governance and Accountability](#), December, at
<http://www.ase.tufts.edu/gdae/Pubs/te/GoodwinMarketFailureFinal2005.pdf> .
- Neva Goodwin, 2008, "From Outer Circle to Center Stage: The maturation of heterodox economics" in *Future Directions in Heterodox Economics*, eds John Harvey and Rob Garnett (University of Michigan Press)
- Albert Hirshman, 1984, "Against Parsimony: Three Easy Ways of Complicating some Categories of Economic Discourse" *AER*, 74, nos 1-2
- Tim Kasser, 2003, *The High Price of Materialism* (MIT Press)
- John Kay, 2013, "Circular Thinking" *RSA Journal* Issue 4
- Alfred Marshall, 1907, "The social possibilities of economic chivalry" in *The Economic Journal*
- Alfred Marshall, 1920, *Principles of Economics*, 8th edition (reprinted by Porcupine Press, Philadelphia, 1961)
- Philip Mirowski, 1989, *More Heat Than Light: Economics as Social Physics, Physics as Nature's Economics*; Cambridge University Press
- Paul Samuelson, *Economics*, first published in 1948, reissued through 19 editions (later editions co-authored with William D. Nordhaus); last edition published in 2009, McGraw–Hill
- Barry Schwartz, 2004, *The Paradox of Choice* (Harper Perennial)
- Tibor Scitovsky, 1976, *The Joyless Economy* (Oxford University Press)
- Amartya Sen, 1977, "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory" in *Philosophy and Public Affairs*, vol 6, pp 317-44
- Adam Smith, 1982, *The Glasgow edition of the Works and Correspondence of Adam Smith* (Oxford University Press)

Richard H. Thaler and Cass R. Sunstein, 2008, *Nudge: Improving Decisions About Health, Wealth, and Happiness* (Yale University Press)

Tim Thornton, 2013, *The possibility of a pluralist economics curriculum in Australian universities: historical forces and contemporary strategies*; Doctoral Thesis, School of Economics, La Trobe University, Bundoora, Victoria 3086, Australia

Amos Tversky and Daniel Kahneman, 1981, "The Framing of Decisions and the Psychology of Choice," *Science* 211(4481):453-458

Author contact: neva.goodwin@tufts.edu

SUGGESTED CITATION:

Neva Goodwin, "The human element in the new economics: a 60-year refresh for economic thinking and teaching", *real-world economics review*, issue no. 68, 21 August 2014, pp. 98-118,
<http://www.paecon.net/PAEReview/issue68/Goodwin68.pdf>

You may post and read comments on this paper at <http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

Placing economists' analyses of antidumping in an antitrust context

John B. Benedetto¹ [USA]

Copyright: John B. Benedetto, 2014
You may post comments on this paper at
<http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

Abstract

Within the legal profession, there has been a long debate about whether antitrust and antidumping are fundamentally related or dissimilar. However, when U.S. economists have written about antidumping over the last several decades, they have frequently understated or even ignored any arguments that ground antidumping in antitrust law and theory. This paper discusses how antidumping law is presented by many modern U.S. economists, and compares those presentations to how the law and legal scholars have viewed the relationship between antidumping and antitrust law. It shows that understanding the relationship between the justifications for antitrust and antidumping laws requires an understanding of the profound historical changes in antitrust law and the perceptions of antitrust law.

This paper finds that most of the modern economic literature on antidumping either (1) does not discuss antitrust when discussing the reasons for antidumping, (2) presents only an incomplete view of antitrust when discussing the reasons for antidumping, or (3) grounds its view of antidumping in a "Chicago" view of antitrust. These factors, while undeveloped in the economics literature on antidumping, are important for understanding the assumptions underlying the modern economic arguments against antidumping law.

- I. Introduction: Antidumping and antitrust, according to modern U.S. economists
- II. A thumbnail history of U.S. antitrust and antidumping Law
- III. The Chicago Way – The reaction against antitrust
- IV. The case against the Chicago view of antitrust
- V. Modern U.S. economists on antidumping
- VI. Non-economists on the relationship between antidumping and antitrust
- VII. Sykes and Morgan – elucidating the debate
- VIII. Ways in which antidumping flows from antitrust
- IX. "Protectionism"
- X. Conclusion

I. Introduction: Antidumping and antitrust, according to modern U.S. economists

The modern U.S. economics literature on antidumping is highly critical of antidumping. Antidumping is described as "broadly negative," "a serious impediment to international trade," and a law that economists "decry."² The criticism seems rooted in assertions that antidumping law is purely "protectionist," with few economically valid justifications.

However, as this paper will show, this critique is based on some crucial assumptions about antitrust law, and the purpose of antitrust law. In their critique of antidumping, modern U.S. economists either (1) do not discuss antitrust when discussing the reasons for antidumping,

¹ Economist with the U.S. International Trade Commission. The views and conclusions expressed in this article are those of the author. They are not the views of the U.S. International Trade Commission as a whole or of any individual Commissioner. The author thanks Harold Brown for his contributions, and several anonymous reviewers for their suggestions. All errors are the author's. Please direct all correspondence to John Benedetto, email: jbner@verizon.net.

² See table 2 below.

(2) present only an incomplete view of antitrust when discussing the reasons for antidumping, or (3) ground their view of antidumping in a “Chicago” view of antitrust.

The language of the first U.S. antidumping law (1916) proceeds from the language of the first U.S. antitrust act (Sherman).³ However, modern U.S. economists, when writing about antidumping law, tend to downplay this connection, or mention it in passing only as if antidumping law, once put in its basic modern form in the Antidumping Act of 1921, has moved completely away from its original antitrust moorings.

This paper looks more closely at the relationship between the economic justifications for antidumping and antitrust laws. It finds that understanding this linkage requires a much closer understanding of the evolution of antitrust law than most modern U.S. economists demonstrate when writing papers about antidumping. Analyzing any linkage also necessitates taking into account the different legal requirements in enforcing an antitrust law domestically versus doing so internationally, where U.S. law holds no sway.

Both antidumping and antitrust law and enforcement have changed since their inceptions. In broad brush, antitrust law has undergone a much more radical revision than antidumping law. When looking at the relationship between the two, one needs to ask: what is meant by “antitrust law”? Do we mean the original understanding of antitrust law, as it was enforced from the New Deal until the 1970s, or do we mean the “Chicago” understanding of antitrust law that has held sway since the 1970s and 1980s? The answer to that question has a profound impact on any economic analysis of whether antidumping law and antitrust law are fundamentally related or at odds.

This paper is not a legal analysis. It is rather a presentation and critique of the economics literature on antidumping, and one that tries to add to that literature by grounding it in an understanding of how economic criticisms of antidumping law need to be analyzed in the context of different understandings of antitrust law.

II. A thumbnail history of U.S. antitrust and antidumping

To understand what the economics literature on antidumping often misses, one must conduct a quick review of developments in the history of antitrust and antidumping. This section and the following sections are not, nor are intended to be, comprehensive legal histories of either antidumping or antitrust law. They are intended to cover broadly the general points of relevance to the economic literature on antidumping. However, some antitrust issues, such as mergers, are not relevant to antidumping and so are not covered at all in this treatment.

There are three main U.S. antitrust laws potentially relevant to antidumping: The Sherman Act (1896), the Clayton Act (1914), and the Robinson-Patman Act (1936). The first U.S. antitrust law, the Sherman Act, criminalized agreements that “intended to operate in restraint of lawful trade.”⁴ However, the Sherman Act was severely limited in its ability to apply to imports because the Supreme Court restricted its use when applied to contracts concluded in foreign countries. In 1916, the first U.S. antidumping law was passed, criminalizing importing

³ Mastel (1998).

⁴ Gifford (1991), quoting the statute.

products below market value with the intent to injure a U.S. industry or restrain competition. The Congressional rhetoric and some of the legal language for this law drew in part on that used for the Sherman Act.⁵

In the 1912 presidential elections, all three parties ran on platforms alleging that the Sherman Antitrust Act had been too lenient on large corporations. After the elections, and viewing the Sherman Act as too restrictive to bring effective prosecutions, Congress enacted the Clayton Act. The Clayton Act, among its other provisions, limits price discrimination when that price discrimination tends to build a monopoly.⁶ Significantly, the Clayton Act does not contain the Sherman Act requirement to prove intent as part of a finding of anticompetitive behavior.⁷

In 1921, Congress passed the 1921 Antidumping Act, the foundation of current antidumping law.⁸ The 1921 Antidumping Act was passed in part because its proponents saw the 1916 law as ineffective.⁹ Just as the Clayton Act does not require intent for establishing that anticompetitive behavior took place, the 1921 Antidumping Act does not contain a requirement of intent to establish that dumping took place. Later U.S. antidumping law (such as Title VII of the Tariff Act of 1930) preserves the basic structure of the 1921 law. Since 1921, U.S. antidumping law has basically defined dumping as international price discrimination that injures a domestic industry.

On the antitrust side, Congress was not finished with expanding antitrust law. Again concerned that the existing antitrust law was not doing enough to help small businesses, in 1936, it passed the Robinson-Patman Act, which forbade price discrimination from the same seller to two different purchasers. The Robinson-Patman Act was aimed mainly at preventing large retailers from putting small retailers out of business by demanding lower prices from suppliers.¹⁰ Nonetheless, the prohibition on price discrimination, extending such prohibitions already present in the Clayton Act, reads somewhat similarly to prohibitions in antidumping law.¹¹

At this point, however, the histories of antitrust law and antidumping law diverge. In the 1950s and 1960s, the enforcement of antidumping law did not result in many affirmative decisions. The Treasury Department determined dumping margins, and often rejected petitions, or took a long time to make a decision, as there was no statutory time limit. In the 1970s and 1980s, however, a series of changes to the antidumping law took the margin decision out of the hands of Treasury and placed it in the Commerce Department. Statutory

⁵ See Finger (1993) and Sykes (1996), as well as Mastel (1998).

⁶ <http://www.antitrustlaws.org/Clayton-Act.html>

⁷ "Although this legislative history suggests that Congress wanted to outlaw predatory pricing and provides some indication of Congress' understanding of predatory pricing, section two of the Clayton Act does not contain an explicit intent requirement. The Act forbids price discrimination whenever discrimination "may" result in a lessening of competition or a tendency towards monopoly. The identification of predatory pricing by the House committee as the focus of the legislation suggests that the Act's drafters largely equated the reduced competition referred to in the Act with predatory pricing. The Act reflects Congress' decision to define the forbidden conduct objectively rather than through a subjective intent requirement." Gifford (1991). Gifford also argues that U.S. Courts did not recognize this lack of intent requirement until 1936, but nonetheless, the law already existed with language not mandating intent in 1914.

⁸ Finger (1993).

⁹ Mastel (1998) and Finger (1993). This point is discussed more below.

¹⁰ http://www.pepperlaw.com/publications_article.aspx?ArticleKey=318

¹¹ See tables 1 and 2 for some examples of scholars who have noted the similarities.

time limits, the right to appeal, and other reforms were also implemented.¹² Also coincident with the first U.S. trade deficits in decades, the number of antidumping (as well as countervailing) cases rose in the early 1980s, although it fell by the mid-1980s and has levelled off since.¹³ However, not all the changes made it easier for U.S. industries to file and win cases. Some changes have made it more difficult to win or keep antidumping margins, e.g., the institution of sunset reviews and restrictions on how Commerce can use available information.¹⁴

On the antitrust side, antitrust law was not vigorously enforced in the Roaring Twenties, as the Progressive Era (which had led to antitrust law in the first place) came to an end. However, in the decades following the New Deal, antitrust enforcement was reinvigorated, reaching the likely apex of its reach in 1962 with the *Brown Shoe* decision and/or in 1967 with the *Utah Pie* decision. These decisions were seen as placing strict limits on price discrimination.

However, The *Brown Shoe* decision, coming only a few years after critics of antitrust law had begun their own attack on the law, was effectively reversed in 1972, as the Courts began to take heed of critics of antitrust law.¹⁵ By 1977, in a major decision (*Brooke*) involving Japanese imports, the Supreme Court was defining the standard for antitrust quite differently than it had before, as described in the following section.¹⁶

In summary, while antitrust and antidumping law flowed from similar economic justifications, their enforcement and legal development followed very different routes. Antitrust enforcement was strongest after the New Deal and until the 1970s. Antidumping enforcement was weakest during this period, and increased in the late 1970s and early 1980s.

The paper now turns to understanding the reasons why antitrust law changed so substantially in the 1970s.

III. The Chicago Way

Beginning in the 1960s, conservative intellectuals, many associated with the University of Chicago, began an attack on the economic justifications for antitrust law.¹⁷ One of these intellectuals was Alan Greenspan (with no connection to the University of Chicago), who in 1962 published the essay "Antitrust." This essay argued that antitrust law punished successful firms for their success, thus harming innovation. "No one will ever know what new products, processes, machines, and cost-saving mergers failed to come into existence, killed by the Sherman Act before they were born. No one can ever compute the price that all of us have paid for that Act which, by inducing less effective use of capital, has kept our standard of

¹² See Mastel (1998) and Prusa (1996).

¹³ U.S. International Trade Commission (2010).

¹⁴ Mastel (1998).

¹⁵ Markovich (2013). See also Page (2008) and Gifford (1991).

¹⁶ See Cann (1996) and Stewart (1996).

¹⁷ The scholars described herein as proponents of the "Chicago" school have different connections to the actual University of Chicago. This paper uses the moniker "Chicago" to mean these critics of antitrust, whether they were professors or students of the University of Chicago.

living lower than would otherwise have been possible.”¹⁸ (Modern analysts differ over whether antitrust law helps or hurts innovation.¹⁹)

A more complete deconstruction of antitrust law was finished by Robert Bork and Richard Posner, both associated with the University of Chicago. Bork, in turn, credited the influence of Chicago’s Economics Department, especially Aaron Director, as influences on his work.²⁰ Bork described many of the terms used in antitrust law, such as “unfair” competition, as undefined. He then examined the Congressional debate over the Sherman Act, and interpreted the intended purpose of the pricing sections of the law as preventing potential harm to consumers, which he stated could only come from predatory pricing.²¹ Later Courts would define predatory pricing as eliminating competitors, continuing to exclude them, and recouping losses from the elimination and exclusion by raising prices later. (Failing to do these three things means that you did not hurt consumers.)²² Bork then concluded that to use antitrust law, one must show harm from predatory pricing, if one is to show any harm to consumers.²³ To this conclusion, later Courts and commentators then added findings from economists that predatory pricing was rare, and even more rarely an effective strategy for the predator.²⁴

Bork’s conclusions remained controversial in the 1960s, but by 1980, had won over the Courts and many policymakers.²⁵ As Judge Douglas Ginsburg described the evolution of Bork’s critique, “[w]hen Bork’s article was first published in 1966, his thesis was novel. By 1977, it had become the conventional wisdom of the federal courts.”²⁶ With several major judicial decisions in the 1970s supporting Bork’s interpretation, and then the election of Ronald Reagan (sympathetic to Bork’s view on the issue), the Justice Department began enforcing antitrust very differently than it had in the 1940s, 50s, and 60s. Enforcement of the Robinson-Patman Act effectively ended. Standards for antitrust law changed to a focus on “consumer welfare.”²⁷

The relevant economic points underlying what has emerged from the Chicago critique of antitrust law are summarized below.

1. Antitrust law is designed to aid consumers; this is the only good economic justification for antitrust law- and the original intent of those who wrote the law.

¹⁸ Greenspan, “Antitrust” (1962).

¹⁹ See Orbach (2010) for a discussion of the debate over whether Bork’s interpretation of antitrust, or his critics’, would be better at encouraging innovation. Additionally, Lynn (2010) presents a detailed argument that an original understanding of antitrust law led to more innovation.

²⁰ Priest (2008), Page (2008), and Markovich (2013).

²¹ See, for example, Lande (1988) and Priest (2008).

²² Cann (1996), Morgan (1996), and Stewart (1996). See also *Matsushita v. Zenith* 475 U.S. 574 (1986).

²³ Lande (1988).

²⁴ Bolton (2000).

²⁵ Nonetheless, Bork’s interpretation of the original intent of the writers of the Sherman Act remains debatable. Writing in 2010, Orbach, who is broadly sympathetic with the “Borkean” changes to antitrust law, states that “all other [non-Bork] studies of the legislative history of the Sherman Act conclude that Bork was too one-sided and misleading in his presentation of the facts.” Orbach (2010).

²⁶ Kirkwood and Lande article, p. 194. (2008). Page describes the shift in the Supreme Court’s view as somewhat slower, and beginning in 1977. See Page (2008).

²⁷ See, for example, Lynn and Longman (2010), Foer and Lande (1999), Lande (1988), and Markovich (2013).

2. The only type of anticompetitive pricing behavior that hurts consumers is predatory pricing. Predatory pricing means eliminating competitors, continuing to exclude them, and recouping losses from the elimination and exclusion by raising prices later. (Failing to do these three things means that you did not hurt consumers.)
3. Predatory pricing is very rare in real life.

All three of these characteristics are important to the debate over antidumping as well. All three are implicitly or explicitly accepted by economist critics of antidumping law and generally form the underpinning (although usually not stated) of their critique of antidumping law.

IV. The case against the Chicago Way

The Chicago view of antitrust has become, for the most part, the way antitrust law is interpreted and enforced since the 1970s. However, it still has its critics, many of whom are grounded in what this paper will refer to as having an “original” understanding of antitrust.

Specific critiques

One critique of the Chicago view, developed by legal scholars Lande and Kirkwood, Orbach, and others,²⁸ points out that Bork misused the phrase “consumer welfare” in his critique of antitrust law. For Bork, “consumer welfare” meant what economists refer to as “total welfare,” i.e., the welfare of both consumers and producers. Bork pointed out how some of the rhetoric around the Sherman Act focused on the harm to consumers, but Kirkwood and Lande respond that such concern for consumers does not mean that the law should focus on total welfare but rather true “consumer welfare,” i.e., the welfare of consumers, exclusive of (rather than inclusive of) the attempted monopolists. In economic terms, for example, consumer welfare could be hurt when consumers lose their rents to monopolists, but under a total welfare standard, this damage would not be counted if the end result were still the same total welfare. Kirkwood and Lande describe Bork as inaccurately using Congressional debate language focused on consumers to justify an antitrust standard of allocative efficiency (i.e., total welfare) rather than a standard of true consumer welfare.²⁹ Thus, Lande and Kirkwood conclude that the history of antitrust does not suggest that Congress’ intention was to allow anticompetitive behavior if it were allocatively efficient.³⁰

Lande and Kirkwood also point out that some of the original rhetoric around the Sherman Act focused on helping small producers that were negatively affected by trusts (i.e. monopolists)

²⁸ See Kirkwood and Lande (2008), Orbach (2010), and Foer (2005).

²⁹ Similarly, Barry Lynn points out that the actual language of antitrust law rarely (once in Clayton and not at all in Sherman) references the word “consumers,” but rather, preventing “unfair” competition. Lynn states that Bork disguised an argument based on efficiency (a word with negative connotations) to consumers (a word with more positive connotations). Unfair competition is not defined as only predatory pricing (allocative efficiency). Lynn posits that regardless of the effect on consumers, certain types of competition are unfair to other suppliers, and such an interpretation is consistent with the way antitrust law is actually written. Lynn (2010).

³⁰ Orbach (2010) and Foer (2005) make related points. Kirkwood and Lande note that if a thief robbed a University of Chicago economist, the law would not ask whether such a robbery was allocatively efficient before penalizing the thief, nor should it. Kirkwood and Lande (2008).

acting as purchasers,³¹ and they also point out how even failed predatory pricing can have negative effects for consumer welfare and for smaller producers. Others have also argued that anticompetitive behavior can do economic damage to consumers without necessarily putting other competitors out of business. For example, predatory pricing may discourage competition without putting competitors out of business, or may discourage innovation and further investment.^{32 33} Still other critiques of the Chicago view of antitrust focus on showing that predatory pricing is really not that rare.³⁴

Many of the above critiques have perspectives that range from still admiring the basic Chicago shift in antitrust law enforcement (Orbach) to agreeing that the primary focus of antitrust law should be lower prices to consumers (Kirkwood and Lande), while acknowledging there is an element of penalizing behavior that targets smaller producers. These critiques offer some potential important analytical comparisons with antidumping law. For example, these papers point out that Congress did not agree to allow anticompetitive behavior that was allocatively efficient, they point out that predatory pricing can have harmful economic effects even if it is not successful (potentially altering what is meant by predatory pricing), and they acknowledge that some of the justifications for antitrust law focused on the welfare of smaller producers relative to the attempted monopolists. As shall be seen, accepting these critiques of the Chicago view of antitrust undermines much of economists' critiques of antidumping law.

Broad critiques

There are also broader critiques of the Chicago view of antitrust, and these broader critiques are also rooted in interpretation of the original antitrust laws and how they were enforced until the 1970s. These broader critiques are based on the idea that the original intent of antitrust law was to prevent concentrated economic power. A supplier that can use pricing power in one market to undercut rivals in other markets can take over those other markets. As it takes over more markets, the supplier can become more influential politically as well as economically.

This point of view was relatively mainstream in the past. In 1944, for example, Henry Wallace warned in the pages of the *New York Times* that “[w]e all know the part that the cartels played in bringing Hitler to power, and the rule the giant German trusts have played in Nazi conquests.”³⁵ More recently, Albert Foer of the American Antitrust Institute has argued that, based on the original writing of antitrust law, the goals of such law should include “decentralization of economic power, freedom, fair distribution of wealth, maintenance of a level playing field, and other ‘public interest’ goals.” Foer also quotes another legal scholar as

³¹ For example, one Congressman noted that four large Chicago livestock buyers suppressed prices paid to ranchers while also raising prices to consumers. Kirkwood and Lande (2008).

³² Foer (2005), Lynn (2010), Bolton et al (2000), Markovich (2013).

³³ Similarly, game theory, much of which developed after Bork's original critique, has shown how seemingly pro-consumer behavior can actually be anti-consumer collusion. Markovich (2013). Nonetheless, such new economic theories may have “qualified” some of the Chicago style decisions, but are still not resulting in more rulings of antitrust violations. Page (2008).

³⁴ Bolton et al (2000).

³⁵ Wallace (1944).

arguing that antitrust law is about ensuring a “level playing field,”³⁶ a phrase often invoked in U.S. trade politics.³⁷

Some of the critics of Chicago antitrust have also warned of the development of larger government without strong antitrust laws. For example, in 1965, Blake and Jones, in a strong critique of Bork’s views on antitrust, described antitrust law as “the chief bulwark against economic and political forces which historically lead first to a monopolized and then to a socialized economy.” Their argument was that a less-concentrated economy is less likely to evolve into one that requires an abundance of government supervision. They are also highly critical of Bork’s emphasis on efficiency as a criterion for antitrust law, because they believe that economic and political diffusion is more important than efficiency. In justifying their interpretation of the reason for antitrust, they quote Senator Sherman’s description of the 1896 law as “a comprehensive charter of economic liberty.”³⁸

Additionally, Barry Lynn points out that the actual language of antitrust law rarely, if ever, references consumers, but rather, preventing unfair competition. Unfair competition was not defined in the legislation as only predatory pricing or allocative inefficiency. Lynn also points out that regardless of the effect on consumers, certain types of competition are unfair to other suppliers, and that such an interpretation is consistent with the way antitrust law is actually written.³⁹

Lynn has also argued that the 1975 Consumer Good Protection Act “legalized price discrimination” and “undid” U.S. antitrust law, which he described as substantially restraining price discrimination prior to this point. To Lynn, this change, along with the changing interpretation of antitrust law in the Courts in the 1970s, contributed to the strengthening of large U.S. retailers’ competitive position relative to U.S. manufacturers, the increased concentration of large U.S. retailers, and ultimately, the concentration of U.S. manufacturing conglomerates (precipitated by the increased concentration in the retail industry) as well.⁴⁰

These broader critiques of the Chicago view of antitrust are likely also consistent with- or at least not contradictory of- justifications for antidumping because they are not grounded in purely predatory pricing, narrowly defined.

This paper now turns to examining how economists have covered antidumping issues, and whether and to what extent they have incorporated the history of antitrust law in their analysis.

³⁶ Foer (2005). Pitofsky also makes a call for the inclusion of these types of “non-economic” goals in antitrust law, although he does not agree with antitrust law helping small producers, while he acknowledges there are extensive quotes from Congressional antitrust debate using exactly such a justification. He simply asserts that the Courts have already reached a compromise not to take such concerns into account anymore. Pitofsky (1979).

³⁷ Similarly, Mitchell argues that when large national suppliers put regional suppliers out of business, less money stays in the local community. Where before local suppliers may spend their money in the community, and generally have held an interest in maintaining a community, larger national suppliers may simply extract profits from the community. To some extent, antitrust law might prevent large national suppliers from using pricing power from other markets to put local suppliers out of business. Mitchell (2006).

³⁸ They also describe U.S. antitrust as making the United States relatively unique among nations of the time, and add that it has allowed the United States to avoid adopting the types of “paternalistic” and even “despotic” regulations employed by the British and Swedes, respectively. Blake and Jones (1965).

³⁹ Lynn (2010).

⁴⁰ Lynn (2010).

V. Modern U.S. economists on antidumping

Economists justifying antidumping

The economic literature on antidumping begins with Nobel Prize winning economist Jacob Viner's "Dumping: A Problem in International Trade" from 1923. Viner defines dumping as price discrimination between domestic and foreign markets, outlines situations in which it may occur, and then analyzes each situation to assess whether it would, on a national welfare basis, cause economic damage to the importing nation. Viner's treatment is balanced and thorough. He notes examples of U.S. dumping in foreign markets and criticizes some (World War I era) U.S. allegations of German dumping. He does not condemn dumping under all circumstances.⁴¹

Nonetheless, Viner does find that dumping is harmful to the importing nation when it involves predatory pricing or when it discourages further investment in the importing nation. However, it is likely that Viner's definition of predatory pricing is not narrow. In defining predatory dumping, he includes not only eliminating competitors but also changing the behavior of competitors and/or weakening them. As for investment, Viner also discusses how sporadic, unpredictable dumping may affect domestic producers' decisions about capital. Thus, Viner's concerns about dumping are not limited to predatory pricing as later Chicago analysts would define it.

In 1977, Wares wrote an analysis of Viner's work on dumping and concluded that, because price discrimination could occur without negative effects on the total welfare of the importing nation, antidumping actions were not justified on as wide a class of cases as some think. However, Wares still concluded that dumping does harm the importing nation in cases when there is attempted foreign monopolization and when there is injury to U.S. producers, albeit with more restrictive conditions than in U.S. law of the time.⁴² Like Viner, Wares attempted a thorough catalog of why dumping may occur and whether it is harmful. He was critical of many of those who believe dumping is a problem. Nonetheless, his work is rarely or never mentioned as a source in any modern economists' works on antidumping, perhaps because he does acknowledge some situations in which dumping may be harmful.

In addition to Viner, another economist writing on the potential reasons for dumping was Ethier in a 1982 paper. Ethier proposes an alternative reason for dumping, i.e., that with demand uncertainty and sluggish factor price adjustments, countries' producers may have motivations to dump in order to keep production lines running when demand disappoints, with a potentially negative effect for the importing nation. He does not conclude that antidumping is either usually or rarely an effective solution in response, but his work demonstrates a possibility in which it might be a welfare-enhancing solution for the importing country.⁴³

⁴¹ Viner (1923).

⁴² Wares (1977).

⁴³ Ethier (1982). The fact that Ethier's theory, like Viner's, comes years after the 1921 law went into effect should not lead one to conclude, as Sykes does regarding Viner's work, that it could not possibly have been a motive for the 1921 law. While economists may have taken a few more years to spell out with logical precision what dumping is and is not, lawmakers may have had the same general idea earlier, even if they had not spelled out all the assumptions and logical boundaries at that time. Additionally, when Hillberry (2011) finds that a demand shock in conjunction with increased imports increases the likelihood of the filing of an antidumping case, he uses that finding to question the motives

However, as will be shown, most modern U.S. economists “decry” antidumping in their articles on the subject.⁴⁴

Irwin as an example of modern U.S. economists

One of the best examples of the general attitude of modern U.S. economists on antidumping can be found in Douglas Irwin’s “The Rise of US Antidumping Activity in Historical Perspective.” Irwin mentions both the 1896 Sherman Act and the 1914 Clayton Act in the opening of his discussion of “the evolution of U.S. antidumping law.” He connects the 1916 antidumping law to the principles of these laws. However, he then describes the 1921 antidumping law as introducing antidumping law “as we currently know it” and describes it as differing “markedly” from the 1916 antidumping law in that:

“The 1916 law focuses on the intent of the exporter, whereas the 1921 law hinges on a finding of price discrimination and injury. ... In the 1916 law, dumping is related to some vague notion of predatory pricing, but in the 1921 law dumping occurs simply if foreign firms charge lower prices in the United States than in their home market, regardless of whether predation is an issue.”⁴⁵

This paragraph shows two points that this paper will develop further. First, many critics of antidumping are possibly also critics of antitrust, but do not criticize antitrust as overtly. Irwin here describes the 1916 law as having “some vague notion of predatory pricing” when it has a similar notion of predatory pricing to that which antitrust law did at that time.⁴⁶ The phrase “some vague notion,” then, could possibly be applied by Irwin to antitrust law as well, and may mean that Irwin shares some or all of the Chicago views of antitrust.

Irwin draws a distinction between the 1916 antidumping law requiring a finding of predatory intent, whereas the 1921 law requires only finding charging lower prices regardless of predation. However, this distinction is not only between the antidumping law of 1921 and the antidumping law of 1916, but also between the 1896 Sherman Act (the basis for the 1916 antidumping law) and the 1914 Clayton Act (in effect when the 1921 antidumping law was passed). Indeed, one might argue that it would not make sense to have only Clayton on the books when one does not have an antidumping law as well to enforce parallel international efforts.

Finally, Irwin does not mention another major reason why the 1916 law was reformulated in 1921. It had proved virtually impossible to prove predatory intent by a foreign firm under no obligation to U.S. law.⁴⁷

of the filing industries. Interestingly, however, Hillberry’s finding might be completely consistent with Ethier’s 1982 paper, which finds that demand shocks in the exporting nation may contribute to dumping.

⁴⁴ Bown (2004).

⁴⁵ Irwin (1997).

⁴⁶ See Mastel (1998) and Finger (1993).

⁴⁷ Mastel (1998). Finger (1993) addresses this issue, and dismisses arguments similar to those of Mastel, but not on the basis of specific Congressional arguments. In some sense, the difference between Finger and Mastel may be rhetorical. Both agree that the 1921 law was designed to rectify perceived problems with the 1916 law. Mastel believes these problems were real, and Finger believes the stated problems were a cover for “protectionism.”

Empirical papers on antidumping motives

Beyond theoretical issues, the recent economics literature on antidumping sometimes asserts that empirically, the motive for most antidumping cases is not really predatory pricing.⁴⁸ When these statements are substantiated, they often cite to one paper, Hyun Ja Shin's 1998 "Possible Instances of Predatory Pricing in Recent U.S. Antidumping Cases." In this paper, Shin concludes that predatory pricing in antidumping cases is generally rare because U.S. and foreign industries are not always concentrated. Unfortunately, his test for concentration involves looking at data on antidumping cases by four-digit SIC code. (The SIC code, for example, has one code for "steel," whereas many U.S. antidumping cases on steel involve much finer distinctions.) This level of aggregation is far wider than the typical product definition in a U.S. antidumping case, and could lead to an underestimate of how concentrated U.S. and foreign industries are.

More broadly, the modern U.S. economics literature on antidumping also consists of a large number of empirical papers that attempt to identify why antidumping cases are filed. Willig asserts that one 1998 selection of these papers (including the aforementioned Shin paper) shows that antidumping is not designed to protect competition from predation in 90 percent of the cases analyzed.⁴⁹ As with Shin's paper, these papers often end up using broad measures, such as wide product categories (much wider than the categories usually used in antidumping cases), to attempt analysis. Hilberry (2011) attempts an analysis at a relatively specific level of product classification, and ends up with more nuanced results than many of the other papers (with broader definitions) do.⁵⁰

Arguably, some of these papers could be examining whether antidumping reflects antitrust motives (such as predatory pricing) or not. However, in the broad class of these papers, the connection with antitrust motives is rarely spelled out. While the papers often contain the economist's censure of antidumping law and/or its results, the papers do not always elucidate whether the economist regards antidumping as an extension of antitrust law.

Other economists have criticized the specifics of how antidumping law is enforced, e.g., the use of "facts available" in determining margins.⁵¹ In these papers, it is not always clear if such critiques are the author's only criticism of antidumping law, or whether the author has other criticisms as well.⁵² In the rare paper (such as Shin's) that addresses specific antitrust-type motives (such as predatory pricing), it is also even more rarely spelled out whether the author has a Chicago or original understanding of antitrust.

In sum, economists' empirical work on the motives for antidumping filings often focuses on broad measures, and does not always explicitly state the authors' beliefs about the underlying

⁴⁸ For example, see Zanardi (2005).

⁴⁹ Willig (1998)

⁵⁰ See a summary of some such papers in Hilberry (2011). As another example, a 2001 paper by Prusa and Skeath, "The Economic and Strategic Motives for Antidumping Laws," concludes that, because nations with antidumping law usually use antidumping law against other nations with antidumping law, antidumping law is employed for "strategic" rather than only "economic" reasons. Prusa and Skeath (2001).

⁵¹ For example, see Moore (2001).

⁵² It is also worth noting that the economics literature frequently discusses lobbying as an explanation for antidumping law, as domestic industries lobby for protection. This type of description interestingly ignores that there are also importing and purchasing firms that lobby against antidumping.

reasons for antidumping, nor how those beliefs connect to the authors' underlying beliefs about antitrust.

The modern U.S. economics literature on antidumping and antitrust

Table 1 summarizes some of the economics literature on antidumping in which either antitrust is mentioned, or in which reasons for antidumping law are elaborated. It is not (nor is meant to be) an exhaustive list of all the literature, but rather a selection of some of the titles that an economist, searching for the latest or most definitive research on the economics of antidumping law, might come across. As can be seen from the table, antitrust law is rarely developed as a motive for antidumping law. If antitrust law is mentioned, the description of antitrust law provided is often brief and may not cover potentially relevant aspects like the absence of an intent requirement in Clayton or the handling of price discrimination in Robinson-Patman. When antitrust law is mentioned, the papers often show a sympathy for a Chicago understanding of antitrust law, without necessarily mentioning other potential understandings.

This paper is not claiming that these papers should always have such a connection clearly drawn. In some cases, it is not the point of the paper, and perhaps in other cases, the authors have thought through the issue but not written it down. Whatever the reason, one can read large swathes of the economic literature on antidumping and not find a detailed discussion of the relationship between antidumping and antitrust law, even when the reasons for antidumping law are discussed. Thus, it is not clear whether the economist in question accepts the basic idea of an antidumping law but questions its enforcement, or whether s/he would prefer to “decry” it altogether, and if so, what the economist thinks about antitrust law.

Table 1 Selected economics papers on antidumping

Name of authors and paper	Summary of relevant portions	Discussion of antitrust
Blonigen and Prusa, "The Cost of Antidumping: The Devil is in the Details" (2003)	This paper describes a "disconnect" between antidumping and competition policy. Describes antidumping law as targeting a vaguely-defined "unfair" practices. States that the most relevant anticompetitive behavior to antidumping is predatory pricing.	Its description of competition policy is based on the Chicago-school understanding, as it claims that antitrust laws do not deem price discrimination as antitrust (without further explanation). It states that "[f]ortunately," Robinson-Patman is no longer enforced. Does not note that antitrust law also uses the word "unfair."
Blonigen and Prusa "Antidumping" (2001)	Reviews the economic literature on antidumping. Asserts that antidumping is "the most serious impediment to international trade."	Draws broad conclusions that antidumping is not about antitrust goal of predatory pricing in one mention.
Bown, "Trade Remedies and WTO Dispute Settlement," (2004)	In a section labelled the "Economic Motivation for Trade Remedy Laws," there is a dismissal of trade remedy as import substitution that generates welfare inefficiencies. He notes that trade remedy laws have been justified by some economists as an insurance or safety valve for nations entering trade liberalization agreements.	There is no mention of antitrust concerns as a potential economic motivation for trade remedy laws, only that economists "decry" antidumping laws.
Galloway, Blonigen, and Flynn "Welfare costs of US antidumping (AD) and countervailing duty laws," (1999)	Models effects of U.S. antidumping laws and finds a net negative effect to the U.S. economy.	Does not mention antitrust. Does not model any benefit to antidumping action other than benefit to producers and government, i.e., the assumption is that the law has no other justification than helping producers and generating tariff revenue.
Mankiw and Swagel, "Antidumping: The Third Rail of Trade Policy," (2005)	Antidumping has a "broadly negative" impact. It was developed to address foreign predatory pricing, but "fair value" goes far astray from only addressing predatory pricing.	Mentions the Sherman Act but no other antitrust laws. States that "By the standards of antitrust, low prices are a problem not when they simply harm other competitors, but when they threaten to wipe out competition and thereby ultimately harm consumers. In practice, this situation is rare."
McGee, "Antidumping Laws as Protectionist Trade Barriers: The Case for Repeal" (1996)	Argues that antidumping laws are purely protectionist and should be repealed.	Discusses the history of AD laws without mentioning antitrust laws. Later, calls predatory pricing the rationale behind AD law. Claims that antitrust law can make the economy less efficient.

Name of authors and paper	Summary of relevant portions	Discussion of antitrust
McGee, "Some Ethical Aspects of Antidumping Laws" (2000)	States that the only reasons for antidumping laws are to stop predatory pricing (which he states "does not exist") or to level the playing field, which contradicts trade theory.	States that there is evidence that antitrust law was designed to "protect producers," and adds that "antitrust laws are being used as weapons to protect less successful competitors from more successful competitors."
Prusa "The Trade Effects of U.S. Antidumping Actions" (1996) and "The Road Most Taken" (1995)	AD law is portrayed entirely as "protection" with industry lobbying to get more.	Does not mention antitrust law in brief backgrounds on antidumping.
Ruhl, "Antidumping in the Aggregate," 2013.	Like Galloway et al above, models the effects of antidumping without any possible positive effects, and adds another potential negative (in this model) effect: importers artificially raising prices.	Concedes that this paper does not model possible predatory pricing, but mentions no other motives for antidumping. Refers only to other modern U.S. economist papers (most cited here) when referencing the reasons for antidumping.
Shin, "Possible Instances of Predatory Pricing in Recent U.S. Antidumping Cases" (1998)	Predatory pricing is generally rare because U.S. and foreign industries- at the wide four-digit SIC level- are not often concentrated. There are no good economic motivations for antidumping other than predatory pricing.	Calls for an antitrust test in antidumping cases, although it should be noted he likely means a "predatory pricing," i.e., Chicago-school test. He acknowledges the Robinson-Patman Act, and states that it has been criticized for similar reasons as he is criticizing antidumping.
Vandenbussche and Zanardi, "What explains the proliferation of antidumping laws?" (2008)	"The only type of 'unfair' dumping, according to economists, is the case of 'predatory dumping.'"	Does not mention antitrust concerns as trade expands as a reason why nations may adopt antidumping laws.
Willig, "The Economic Effects of Antidumping Policy" (1998)	Divides dumping into nonmonopolizing and monopolizing categories, the latter of which includes both strategic dumping and predatory pricing, and could have negative effects on the importing market.	Notes that antitrust policy is focused on predation, and that predation requires similar criteria to those described above in this paper under antitrust. Concludes that the empirical record shows that antidumping is usually not undertaken to protect competition.
Zanardi, "Antidumping, a Problem in International Trade," (2005)	As predatory dumping is rare, antidumping imposes a larger consumer cost on society than benefits producers. Predatory dumping is the only "sound economic" motive for AD.	Other than brief discussion of predatory pricing, there is only a brief mention of antitrust in a footnote.

Source: Author's summary of listed papers. See bibliography for complete listing.

To summarize the modern U.S. economist's theoretical case against antidumping as an analog to antitrust, the following points are usually present, although they are not often stated, or if stated, often stated very quickly and with little elaboration.

1. The 1916 antidumping law may have been influenced by the 1896 antitrust act, but there is little to no acknowledgment of later antitrust developments (such as the

removal of the requirement of intent in Clayton and the emphasis on price discrimination in Clayton and Robinson-Patman) that are similar to antidumping law.

2. Predatory pricing is almost the only, or the only, reason for antitrust law, and as such, should be the only motivation for antidumping law.
3. Predatory pricing is rare.

As previous sections showed, none of these three points is beyond debate.

The Stiglitz framework

Economist Joseph Stiglitz has also offered a critique of antidumping in his 1997 paper “Dumping on Free Trade.”⁵³ While the paper uses most of the typical modern U.S. economist’s unstated assumptions about antidumping and antitrust, it also offers an interesting framework with which one can consider antidumping in a more balanced fashion.

Stiglitz begins by noting that, from a static perspective, dumping would seem to make the importing nation better off. However, he notes two exceptions to that possibility: dumping due to predatory pricing; and “new trade theory” effects from dumping. While Stiglitz does not elaborate in great detail on what he means by potential new trade theory effects, he may include investment effects from having market power in one country, and/or from externalities. If so, these “new trade theory” effects may overlap with the type of investment effects that Viner referred to in 1923.

Stiglitz continues on, though, to an analysis more typical of modern U.S. economists. He asserts that “since the 1921 Antidumping Act, it has not been necessary to demonstrate predatory intent or effect, so any international price discrimination... has been proscribed.” In doing so, he does not note that since the Clayton Act of 1914, intent has not always been necessary in antitrust either.

Stiglitz then states that “since the 1974 trade act, the focus has shifted from preventing price discrimination to preventing sales below fully allocated average cost.” He goes on to claim there are reasonable economic reasons such sales may occur, citing two papers.⁵⁴ He does not make any analysis of whether such reasons may be more or less likely than anticompetitive reasons for dumping, nor does he take into account the difficulties for U.S. government in determining whether foreign companies are engaged in international price discrimination for reasonable or for anticompetitive reasons.⁵⁵ As will be shown below, supporters of antidumping law have defended the reliance on average cost.

Stiglitz then offers a useful method of considering the pros and cons of antidumping. Borrowing from statistics, he describes antidumping policy as potentially leading to either type I errors (not using antidumping law when it should be used) or type II errors (using

⁵³ Stiglitz (1997).

⁵⁴ As shall be seen below, some U.S. legal writers, like Stewart, do not agree that long-term sales below fully allocated average cost can reflect anything other than anticompetitive factors. Stiglitz does not address whether such sales below fully allocated average cost are long-term or not.

⁵⁵ Stiglitz goes on to critique particular methods used in calculating antidumping margins, some of which may have changed since he wrote the paper. This treatment is not meant as a complete summary of his entire paper.

antidumping law when it should not be used). This framework can be seen as recasting the antidumping debate into an empirical question in which antidumping supporters view type II errors as less frequent than type I errors, and antidumping critics hold the opposite view. (Stiglitz clearly falls into the antidumping critics category, but interestingly does not call for abolishing antidumping, but rather restraining it to try to reduce type II errors).

This framework is similar to the debates over antitrust, in which the Chicago critics have adopted a restricted set of reasons for when antitrust law should be used (i.e., only in cases of predatory pricing) and operate under a presumption that predatory pricing is a relatively rare phenomenon. On the other hand, original antitrust supporters hold a view that antitrust is about more than just predatory pricing, narrowly defined, and such anticompetitive behavior is not necessarily rare.

Economist Robert Willig reached a similar, relatively balanced conclusion in 1998.⁵⁶ He describes various categories of dumping, and concludes that, at least in theory, some of these categories can be negative for total welfare in the importing nation and sometimes globally. However, his paper, which introduces a series of papers that attack antidumping (see table 1), concludes that the empirical record is that antidumping is used to distort rather than encourage competition. There are serious potential difficulties with the techniques used in these (and other) papers, but Willig's major contribution is again reiterating the Viner finding that there are circumstances in which dumping can be harmful, including not only predatory pricing but also "strategic dumping."

As can be seen in table 1, however, the relative balance of Stiglitz and Willig stands in contrast to most of the modern U.S. economists who write about antidumping. Most such works are more likely to be filled with the warnings about increasing use of antidumping, descriptions of antidumping as nearly pure "protectionism," and little mention of the wider context of antitrust law and history.

VI. Non-economists on antidumping and antitrust

In contrast to modern U.S. economists, non-economists (mostly legal scholars) have offered a wider range of analyses both on whether antidumping law is a desirable policy, and whether antitrust and antidumping are fundamentally related or not. A selection of these analyses is summarized in table 2. Not all of these papers were written specifically to address the relationship between antitrust and antidumping, but by touching on the issue (or not touching on it when it could be relevant) show the authors' underlying assumptions or analyses of the relationship.

While again this table is not a complete presentation of every paper written on the reasons for antidumping, it does show that in the selection presented, authors coming to a similar conclusion as U.S. economists (i.e., that antidumping and antitrust are not closely related) do so with similar underlying assumptions, i.e., little discussion of Clayton or Robinson-Patman, dismissal of Robinson-Patman when it is brought up, and dismissal of predatory pricing as rare.

⁵⁶ Willig (1998).

Table 2 Selected non-economic writings on antidumping and antitrust

Name of authors and paper	Summary of relevant portions	Discussion of antitrust
<i>Papers with a similar perspective to modern U.S. economists</i>		
Applebaum, "The Interface of Trade/Competition Law and Policy: An Antitrust Perspective," (1987)	Antidumping law does not have the same standards as antitrust law, including those of intent and ability to monopolize successfully. Antidumping cases could even allow antitrust violations when filers act in concert.	Concludes that there are fundamental differences between antitrust and antidumping law.
Cann "Internationalizing Our Views Toward Recoupment and Market Power: The Antidumping/Antitrust Dichotomy through WTO-Consistent Global Welfare Theory," (1996)	"Antidumping laws are protectionist measures applied regardless of market structure, consumer welfare, or the relative efficiencies of foreign and domestic industries. They effectively outlaw international price discrimination, while ignoring its domestic equivalent, by treating predatory and nonpredatory price levels similarly."	Antidumping and antitrust law are in "dichotomy" because antidumping law is broader, with some examples being based on recent Court decisions restricting antitrust findings (such as <i>Brooke</i>). However, he also critiques the Chicago view of antitrust, arguing that its narrow definition of predatory pricing has led to firms filing antidumping cases because it is so difficult to win antitrust cases on foreign predatory pricing. The author proposes a "antitrust-based predatory pricing remedy to be considered during future WTO negotiations."
Finger, "Unfair Trade and the Rules on Dumping," in <i>The World Trading System</i> (1989)	While acknowledging early antidumping law as an extension of antitrust, asserts that later antidumping laws and changes to laws were entirely protectionist.	Does not discuss the Clayton Act or the Robinson-Patman Act, even as he focuses on the lack of intent in antidumping.
Hoekman and Mavroidis, "Dumping, Antidumping and Antitrust," (1996)	Antidumping law should include consideration of consumer welfare, and assessments of whether the foreign market is contestable. Proposes various ways of making antidumping law more like modern antitrust law.	While antidumping proponents portray antitrust as the motivation for antitrust, the relationship is "no longer the case." Most economists "agree that predatory dumping is the exception, not the rule. Proponents of antidumping are concerned, implicitly if not explicitly, with the continued existence of national firms that produce a good." These statements are not developed.
Jackson, <i>The World Trading System</i> (1989)	While connecting the 1916 antidumping law to antitrust, does not connect the 1921 law to antitrust. Offers reasons why firms may price discriminate between nations while not dumping.	Does briefly acknowledge some similarities between antidumping and Robinson-Patman, but dismisses Robinson-Patman as criticized by economists and no longer enforced.
Victor, A Paul. "Antidumping and Antitrust: Can the Inconsistencies Be Resolved?" (1982)	The 1916 Antidumping Law was modified because it was "virtually unenforceable." However, successor laws have not distinguished between the goals of stopping anticompetitive behavior as opposed to protecting domestic industries.	Antidumping's material injury standard is different than Robinson-Patman's competitive effects standard. Antidumping law has been enforced without the traditional antitrust objective of increased price competition.

Name of authors and paper	Summary of relevant portions	Discussion of antitrust
Wu, "Antidumping in Asia's Emerging Giants," (2011)	Discusses antidumping in Asia, with lots of references to protectionism "rearing its ugly head," antidumping being "backdoor protectionism," etc.	There are two mentions of antitrust, and no detailed discussion of antitrust. The author thanks many economists for their contributions.
Papers with other analyses		
Anthony, "The American Response to Dumping from Capitalist and Socialist Economies," (1969)	Dumping "may impair normal conditions of competition" and has potentially ill effects in both the long and short run.	Does not discuss much the relationship to antitrust law, other than to footnote that others have commented on the issue.
Eckes (historian and government official), "The Interface of Antitrust and Trade Laws- Conflict or Harmony?" (1987)	Trade remedy laws emerged from similar concerns as antitrust. Dumping and subsidies lead to distorted international resource allocations.	Trade remedy laws are compatible with antitrust laws and the goals of antitrust laws (increasing competition), although they may conflict in the short run.
Epstein, "The Illusory Conflict Between Antidumping and Antitrust," (1973)	Concludes that the only way protracted dumping can take place is with market restraints in the exporting market.	States that "antidumping regulations act as an extension of antitrust legislation," allowing regulation of foreign anticompetitive behavior when the effectiveness of U.S. antitrust law is limited (i.e., in foreign countries).
Garten (government official), "New Challenges in the World Economy: The Antidumping Law and U.S. Trade Policy," (1994)	Dumping results from closed foreign markets, anticompetitive markets, and/or nonmarket conditions in exporting country. Dumping " sends false signals to the market.... causes resources to be misallocated ... has a dramatic effect on investors' decisions." The belief that lower prices will result from allowing dumping is "shortsighted" and such lower prices do "not reflect genuine free competition."	Antitrust law was an important distinction between U.S. law and that of most countries. The emergence of a culture of antitrust, combined with foreign dumping, led to the extension of antitrust principles into antidumping.
Gifford, "Rethinking the Relationship between Antidumping and Antitrust Laws," (1991)	Traces the history of antitrust law, for example, the removal of intent with the Clayton Act, the use of the Robinson-Patman Act to make price discrimination illegal, and the Chicago reaction to antitrust law.	"Despite superficial inconsistencies, however, it appears possible to reconcile trade and antitrust laws to a substantial extent." He describes the <i>Utah Pie</i> antitrust case (which provoked strong reaction from the Chicago-style critics) as similar to an antidumping case (while not approving of <i>Utah Pie</i>).
Mastel, <i>Antidumping Laws and the U.S. Economy</i> (1998)	Provides an economic rationale for antidumping law as providing a counterweight to dumping based on sanctuary markets, subsidized industries, and/or nonmarket economies; questions Finger's interpretation of antidumping history.	Attributes the development of the 1921 antidumping law to practical problems with enforcement of the 1916 law that had been modeled more directly on antitrust. Describes antitrust and antidumping as different in that antidumping includes more than predatory pricing, "narrowly defined."
Stewart, "Why Anti-Dumping Laws Need Not Be Cloned After Competition Laws Nor Replaced By Such Laws," (1996)	Justifies antidumping as a response to foreign firms that are selling below all of its costs, as doing so implies that such firms "be in a nonsustainable position, receiving subsidies, or cross-subsidizing its losses with supra-competitive prices elsewhere" and will send false signals to the market.	Notes that there are no internationally-agreed competition laws, and so antidumping law allows better resource allocation than antidumping law. Also criticizes more recent developments in antitrust law, such as the focus on only average variable costs (and not all costs), and more recent unwillingness to find antitrust when there is cross-subsidization. Criticizes the <i>Brooke</i> decision.

Name of authors and paper	Summary of relevant portions	Discussion of antitrust
United States Permanent Mission to the World Trade Organization (WTO), "Observations on the Distinctions between Competition Laws and Antidumping Rules," 1998.	Argues that antidumping deals with "different" concerns than competition policy. Antidumping law is grounded in the need to address remaining "market distortions" that would otherwise do damage in the world trading system.	While stating that antidumping and antitrust are different, the paper also notes that "[t]here is an important linkage" between antidumping and antitrust laws, in that one cause of dumping can be "the absence of, or the lack of adequate enforcement of, meaningful competition laws."
Wood, "'Unfair' Trade Injury: A Competition-Based Approach," (1988)	Proceeds from the assumption that there can be legitimate cases of injury from "unfair" imports, when such imports are lower-priced than domestic product due to artificial advantages.	To avoid protecting domestic monopolies, antidumping law could be brought into alignment with antitrust law by having an injury test that determines whether the U.S. industry is also benefitting from a restricted domestic market.

Source: Author's summary of listed papers. See bibliography for complete listing.

However, as can also be seen in table 2, writers that connect antidumping and antitrust often do so by appealing to a broader understanding of antitrust than just predatory pricing (e.g., Garten, Stewart, Mastel) and/or as an extension of predatory pricing policy with an eye to enforcement of such laws on foreign firms (Epstein, Mastel).

Several legal authors that criticize antidumping for its differences from antitrust (including Cann and Wood in table 2) propose replacing current antidumping law with more of an international antitrust standard. Their critiques of antidumping, although grounded in a Chicago view of antitrust, also are not as rhetorically severe as those of some of the economists in table 1. They do not necessarily propose that antitrust is a law rarely needed, so that invocation of it is likely just a justification for protectionism. (Indeed, Cann argues that the difficulty in winning legitimate antitrust cases has encouraged some firms to file antidumping cases instead.⁵⁷) Similarly, the 1998 U.S. communication to the WTO (see table 2 above), while arguing against folding antidumping law into competition law, notes that one cause of dumping can be the lack of competition law, or enforcement thereof, in other countries.

These proposals for an international antitrust standard to replace antidumping show some important parallels between antitrust and antidumping law. Nonetheless, implementing such proposals raises many questions, such as the issue of defining what is meant by antitrust violations and whether that includes everything that is meant by dumping. Such proposals also need to address the issue of whether other countries have an interest in reconciling their antitrust laws with U.S. antitrust law. Will other countries' governments regularly cooperate with antitrust actions filed by U.S. firms?

VII. Sykes and Morgan – elucidating the debate

Two legal papers which offer particular insight into the issue of the relationship between antitrust and antidumping are Alan Sykes' "Antidumping and Antitrust: What Problems Does

⁵⁷ Cann (1996).

Each Address?,”⁵⁸ which argues that antitrust and antidumping address fundamentally different issues, and Clarisse Morgan’s “Competition Policy and Anti-Dumping Is it Time for a Reality Check,”⁵⁹ which argues that antitrust and antidumping address fundamentally the same issues, but from a domestic versus foreign perspective. Understanding these two papers can illuminate the antitrust context for economics papers on antidumping.

Sykes

On the surface, Sykes’ paper is typical of many of the legal and economic critics of antidumping in its view of antitrust. The paper is highly critical of antitrust, especially the “misguided” Clayton Act and the “protection” motives of the Robinson-Patman Act.

However, Sykes also shows he understands the fundamental similarities in the justifications for antidumping and antitrust as well. Sykes’ paper contains many fascinating quotes from the original authors of both antitrust and antidumping law. He quotes Senator Sherman (of the Sherman Act) noting that trusts may both raise or lower prices. He also quotes Congressional sources that justified the Robinson-Patman Act as helping consumers in the long run by ending unfair business practices. He quotes Judge Learned Hand in the Alcoa antitrust case as saying that Congress “did not condone ‘good’ trusts and condemn ‘bad’ ones; it forbade all. Moreover, in so doing, it was not necessarily actuated by economic motives alone.” Sykes also quotes Supreme Court Justice Earl Warren as stating that the Court “cannot fail to recognize Congress’ desire to promote competition through the protection of viable, small, locally owned businesses” in describing antitrust law. These references certainly sound similar to, or at least consistent with, the justifications for antidumping law.

However, in his paper, Sykes advances two separate notions, the latter of which is at odds with those references. First, Sykes states that with Bork’s critique of antitrust law, “[a]s enlightened economic thinking about antitrust policy developed, the judges could thus embrace and operationalize economic thinking in the law.” This statement is another way of saying that the Bork critique allowed antitrust law to be changed, by the Courts and through enforcement, to reflect the findings of modern U.S. economists, especially those more critical of older antitrust understandings. Stripped of its rhetoric justifying the changes to antitrust law, Sykes’ statement is correct in noting that antitrust law enforcement has changed, dramatically, from its early years, based in part on Bork’s (controversial)⁶⁰ reading of the intent of the writers of the Sherman Act.

Most importantly, though, Sykes goes on to argue that the language used to justify antidumping law is too vague, and too different from the language of antitrust law to allow such a change. He states that the multiple references to predation in the language of those proposing antidumping law were “largely pretense,” but his justification for this characterization hinges on an interpretation of antitrust as being about predatory pricing only and on his own characterization of how the original bill passed.

Sykes writes that “[o]riginally marketed as antipredation measures, [antidumping laws] are now written in a way” that compels authorities to impose duties whenever there is injury from

⁵⁸ Sykes (1998).

⁵⁹ Morgan (1996).

⁶⁰ Sykes acknowledges that Bork’s reading of Sherman’s intent was “controversial.”

lower cost imports, a broader class of cases than predation. However, Sykes omits noting that the original concept of antitrust did not rule out a broader class of cases than predation, as narrowly defined by later economists. Sykes also does not note any difficulty that enforcement authorities might have with enforcing antipredation laws overseas, and whether that might necessitate a different, albeit second-best, standard, as perhaps illustrated by the shift from the 1916 antidumping law to the 1921 antidumping law.

Second, while Sykes does acknowledge that Viner did not justify antidumping only on the grounds of antipredation, he argues that because Viner's book was not written until 1923, two years after the 1921 law, then the writers of the law could not have had such other, nonpredation motives in mind. This logical step is flawed. While modern economists and competition lawyers may have a very specific idea of "predation" (involving lowering prices to force the exit of competitors, taking their market share, and then raising prices while excluding new entrants), it is not clear that the law's Congressional writers would have had such a restricted definition, and very likely might have included negative investment effects (of the kind referenced by Viner, or of the effects of failed predation or incomplete predation) as an example of "predation."⁶¹

At a broad level, Sykes' paper is typical of many of the legal and economic analyses that do not see much relationship between antitrust and antidumping. He implicitly assumes the Chicago understanding of antitrust is the correct one, and bases his analysis off of that assumption. However, Sykes goes beyond most of his intellectual comrades in delving deeper into the relationship between antitrust and antidumping. He acknowledges the clear linkages between the two types of law and tries to see why they have diverged.

Morgan

Clarisse Morgan has also offered an important, if overlooked in the economics literature, paper showing some of the essential conceptual similarities between antitrust and antidumping.

She opens her paper by stating that the price discrimination standard in antidumping and antitrust is different, but again, her argument is based on an implicit Chicago definition of antitrust. For example, she states that in antidumping cases, the first issue is whether import prices are below normal value, but in antitrust, the issue is whether price discrimination "has the possibility of driving a competitor or competitors in the lower-priced market out of business, and then keeping that market closed to competition, or 'noncontestable.'" She cites the 1986 antitrust case *Matsushita* as an example of a Court finding no predatory intent by Japanese exporters even though the Court did find that the Japanese industry had a cartel-like structure and did sell into the U.S. market at a lower price than in the Japanese market. The Court did so because it argued that the U.S. industry had survived for many years despite the presence of the Japanese imports.⁶²

Morgan does not note, though, that the "noncontestable" criteria in antitrust is relatively new, and related to the idea that antitrust violations must involve predatory pricing, "narrowly

⁶¹ Additionally, Sykes first describes the motivations for antidumping law as being too vague, and then describes them as specific enough to exclude the non-predatory motives listed by Viner. These descriptions by Sykes may not be logically consistent.

⁶² See also *Matsushita v. Zenith* 475 U.S. 574 (1986).

defined” (Morgan).⁶³ It is not clear that earlier understandings of antitrust hinged on such a narrow definition. The *Matsushita* case was a 5-4 decision, and possibly earlier Courts, with an older understanding of antitrust, might not have reached the same decision. In other words, while Morgan is using a Chicago view of antitrust to reach her findings, her findings would likely be bolstered by using the original view of antitrust.

Morgan goes on to argue that there is a “broad middle ground” between completely open, contestable foreign markets and completely predatory ones. This broad middle ground might include, for example, markets in which the government offers difficult-to-detect assistance to an industry or does not enforce competition laws. She points out that critics of antidumping ignore this broad class of possibilities when asserting that all non-predatory price discrimination is not harmful.⁶⁴

Morgan’s above points are an important counterweight to critics of antidumping, hitting on the nub of their basic assumption that in general, international price discrimination is probably not predatory, and so antidumping is very different than antitrust. Once again, the key issue of exactly what is meant by antitrust is key to understanding whether antidumping and antitrust are similar or not.

From price discrimination, Morgan goes on to consider the case of selling below cost. She argues that in general, in both domestic and foreign contexts, allowing selling below average variable cost for anything other than some short-run sales is likely not economically optimal.⁶⁵ She also points out that “It is a substantial theoretical leap, which seems to go largely unquestioned, to move from the statement that pricing at variable cost is not necessarily inconsistent with economic efficiency, to the conclusion that pricing at variable cost proves that a firm is operating in a nonpredatory, economically efficient manner.” (The latter would be a standard more likely to meet Chicago antitrust standards, but, as shown earlier in this paper, would perhaps not pass muster with older understandings.)

Morgan argues instead to use an antitrust standard of predation as “as pricing that yields profits below the opportunity cost of money, i.e. below the profitability of competing investments.” This definition, she then points out, would yield an analysis very similar to the analysis in current antidumping law.⁶⁶

⁶³ Interestingly, another defender of antidumping, Greg Mastel, also finds that antidumping covers a wider class of problems than antitrust, but again is implicitly using the Chicago definition of antitrust. Mastel (1998).

⁶⁴ Morgan also points out that in a general equilibrium framework, when one country’s industry dumps from protected home markets at lower than the world price, it distorts investment decisions and causes net welfare losses.

⁶⁵ As Mastel and Stewart also point out, in the long run, average costs are the same as average variable costs.

⁶⁶ She points out that the injury standard would still be different, in that antitrust law imposes a “recoupment test” testing for the impact on consumer welfare, while antidumping law imposes an injury test that looks at the impact on producer welfare. She argues that in cases in which dumping is causing a negative resource allocation, it is not always true that a consumer welfare test based on recoupment will be more societally optimal than a producer injury test. In addition to Morgan’s arguments, the modern antitrust understanding of the recoupment test—that the predator must be able to recoup all its losses from low pricing—may not be consistent with an original understanding of antitrust law. Thus, under an original understanding of antitrust law, the arguments for an antidumping law might not even require this argument.

Morgan concludes that “once the dumping and predation frameworks are reduced to their essential forms, they can be seen in some cases to resemble each other rather closely.”⁶⁷ It is worth noting once more that in her analysis, any remaining differences are likely based on a Chicago school understanding of antitrust law that requires effective predatory pricing in order to find an antitrust violation. As developed earlier in this paper, an older understanding of antitrust law would thus likely be even more similar to current antidumping law.

Also interesting to note in Morgan’s work is that she acknowledges some of the potential pitfalls of antidumping (e.g., creating anticompetitive opportunities for U.S. producers and penalizing price discrimination that is not anticompetitive), but states these outcomes are likely rare. Her analysis is, then, somewhat consistent with that of Stiglitz (in his Type I and Type II error framework), who comes to the opposite conclusion about which type of error is more likely. Their arguments suggest that one of the main differences over antidumping is a question of how likely anticompetitive behavior is, and again, this question dovetails with the issues underlying antitrust. Chicago antitrust believers tend to believe that predatory behavior is rare and minimally harmful; original antitrust believers had (and their modern analogs have) a wider view of the potentially negative effects of anticompetitive business behavior.

VIII. Ways in which antidumping flows from antitrust

With this history of antitrust in mind, one can now turn to ways that the economics literature on antidumping can be better understood by more consideration of the parallels with antitrust law. The following section discusses four ways in which antidumping law is consistent with the original understanding of antitrust law, and even flows from it.

First, antidumping takes place in a world in which the normal methods of proving antitrust violations, e.g., subpoenaing the records of accused antitrust violators, are not easy or possible. Foreign firms are under no obligation to obey U.S. antitrust laws, nor subject to penalties by U.S. authorities. Thus, it makes sense that antidumping law may have different methods than antitrust. For example, proving intent or predation might become something more like proving injury due to price discrimination or below-cost pricing.

Predatory pricing might be more likely to exist in an international context because of protected home markets, a phenomenon less likely to exist within a nation, especially one with tough antitrust laws (like the United States after the New Deal). Another country may have no antitrust law or weaker antitrust law, resulting in some of its firms having market power within that country, and allowing them to dump into the U.S. market. Similarly, other markets may be closed to international competition, for reasons that are not easily discernible, such as close (but not public) relationships between firms or between firms and the government of that nation. Again, in such circumstances, the firms of that country may dump into the U.S. market.⁶⁸

⁶⁷ Her analysis here can also be seen as a refutation of the arguments presented by Applebaum and Victor in table 2.

⁶⁸ Similar arguments were made by the U.S. mission to the World Trade Organization in 1998, when the United States argued that antidumping law covered a wider class of policy issues than competition policy does. See United States Permanent Mission to the World Trade Organization, “Observations on the Distinctions between Competition Laws and Antidumping Rules,” 1998.

Within the U.S. market, such dumping would be covered by antitrust law, and indeed, likely be an antitrust violation, with all the firms concerned being subject to U.S. antitrust law. However, when international trade is considered, the U.S. market will now be open to products produced in areas outside the scope of U.S. antitrust law. Attempting to close that loophole is arguably consistent with antitrust law.

Second, as has been shown, some U.S. antitrust law, notably the Clayton and Robinson-Patman Acts, do not require predatory intent and do, under some interpretations, penalize price discrimination. The debate over how to interpret these laws was a large and expansive one in the history of antitrust, and is still ongoing. One's view of antidumping law may depend heavily on one's view of this debate in antitrust. It also likely depends on one's views of the related issues of what is meant by predatory pricing, and how often predatory pricing occurs. The economics literature on antidumping often reflects an unstated acceptance of the Chicago view on these antitrust issues.

Third, whether predatory pricing is the only justification for antitrust (the Chicago view) or not, there may be additional justifications for a different type of competition policy in an international trade context. Allowing dumping to take place can affect domestic firms' investment decisions, as noted as far back as Viner and more recently. That is, otherwise economically efficient U.S. producers might not make investments in U.S. production if the market in which they produce is subject to dumping. The economics literature on antidumping rarely mentions this possibility as a justification for antidumping law.

Fourth, the economics literature almost never mentions countervailing duty law, which is designed to counter explicit foreign subsidies. Countervailing duty law follows a very similar legal path to antidumping law in the United States, with investigations needing to show injury in the same way but by reason of subsidy (rather than dumping) in the foreign market.

John Maynard Keynes argued for countries to have the flexibility to subsidize their industries when necessary as part of the international economic system, but then for their trading partners to respond with countervailing duties.⁶⁹ Keynes clearly saw such a system (subsidies with countervailing duties) as consistent with an international rules-based trading system. A similar argument could be constructed for antidumping law, in which countries would be allowed flexibility for their own competition policies, and other nations could react with antidumping law.

IX. "Protectionism"

The modern economics literature on antidumping frequently describes antidumping as "protectionist." One difficulty with the logical argument against antidumping being related to antitrust is the imprecise use of the word "protectionism." Interestingly, many of these economic authors sometimes refer to some types of antitrust, such as Robinson-Patman, as "protectionism." Clearly, if protectionism is defined traditionally as placing barriers to international trade, antitrust law cannot be protectionist when it involves relations between domestic firms.

⁶⁹ Horsefield (1969).

What those who use “protectionist” to describe antitrust law likely mean is summarized in the Supreme Court’s 1993 (i.e., post-Chicago antitrust) opinion that U.S. competition law should “protect competition and not competitors.” Thus, “protectionism” is probably defined as meaning “protecting competitors.” Nonetheless, this definition is still very vague. Exactly how does one protect competition without protecting competitors? Any sanction on one competitor could be portrayed as “protecting” another competitor. There is a danger in some of this literature that “protectionism” can simply be defined as something that the specific author does not like. At best, one could say that “protectionism” has been unofficially redefined to mean any competition policy that goes beyond protecting against predatory pricing, as defined by Bork et al. However, such an explicit definition is not always elucidated.

Antidumping critics may argue that antidumping actually creates problems for consumers, by protecting inefficient or oligopolistic domestic firms. Whether this argument is correct or not, it has a near-exact parallel to the critics of antitrust from the mid-20th century, when scholars like Robert Bork argued that antitrust was “protectionism” for smaller, less-efficient producers. This argument again shows the fundamental parallel between the motivations for antitrust and antidumping, and the motivations for criticism of them.

X. Conclusion

Legal scholar William Page has described the underlying tension between the Chicago vision of antitrust and the original view as a tension between “evolutionary” and “intentional” visions of the market. In the evolutionary vision, the market is a self-regulating mechanism that allows free exchange among individuals, none of whom has lasting influence on the market. On the other hand, the “intentional” vision sees the market as tending toward monopoly, and resulting in coercion on individuals and smaller businesses.⁷⁰

Page’s framework is useful to show the broad similarities between views of antitrust and antidumping. Modern U.S. economists are perhaps more likely than legal scholars, historians, and even economists of previous years to view relatively laissez-faire economies as tending toward the most efficient equilibrium. Thus, their literature on antidumping will reflect a point of view similar to that of the Chicago critics of antitrust, i.e. that antidumping (and antitrust) address a problem that is rare, and the cure may be more damaging than the problem. It is important to understand this fundamental, and often unstated, belief of modern U.S. economists when reading their analysis of antidumping.

Though antitrust law and antidumping law are not exactly the same, they proceed from many of the same concerns, i.e., that of whether government ought to restrain sellers with market power from hurting other sellers through low pricing based on that market power, and whether such actions also hurt consumers. This paper has shown that if antitrust law is interpreted as Robert Bork and the other “Chicago”-style critics argued (successfully to date), then antidumping law may be interpreted as broader than antitrust law as it is not motivated entirely by predatory pricing and its effect on consumer welfare. Nonetheless, an argument for antidumping law as currently used might still be tenable as a method of dealing with the inability of the U.S. government to obtain information abroad as it would in an antitrust case. However, if antitrust law is interpreted as it was in the 1940s through the 1960s, then

⁷⁰ Page (2008).

antidumping law is more likely to be at least a rough analog of antitrust law applied to foreign trade.

Perhaps the best way to understand the debate over antidumping law is to view the world in the way Stiglitz and Morgan do, i.e., in a framework in which imperfect information is going to lead to errors either way. Either antidumping cases will be filed when they should not, or cases will not be filed when they should. Stiglitz and Morgan both raise this broad framework, and yet reach different conclusions about the efficacy of antidumping law. No matter what, understanding the analytical frameworks of economists who discuss antidumping requires also understanding what they believe about the nature and goals of antitrust.

References

<http://www.antitrustlaws.org/Clayton-Act.html> downloaded December 16, 2013.

Applebaum, Harvey M. "The Interface of Trade/Competition Law and Policy: An Antitrust Perspective," 56 *Antitrust Law Journal* 409 1987.

Anthony, R.A. "The American Response to Dumping from Capitalist and Socialist Economies," *Cornell Law Review*, Vol. 54 No. 2, January 1969.

Blake, Harlan M., and William K. Jones. "In Defense of Antitrust," 65 *Columbia Law Review* 377, 1965.

Blonigen, Bruce A., and Thomas J. Prusa. "The Cost of Antidumping: the Devil is in the Details," *Policy Reform*, December 2003.

Bolton, Patrick, Joseph F. Brodley, and Michael H. Riordan. "Predatory Pricing: Strategic Theory and Legal Policy," *Georgetown Law Review*, August 2000.

Cann, Wesley A. "Internationalizing Our Views Toward Recoupment and Market Power: The Antidumping/ Antitrust Dichotomy through WTO-Consistent Global Welfare Theory," *University of Pennsylvania Journal of International Economic Law*, 1996.

Eckes, Alfred E. "The Interface of Antitrust and Trade Laws- Conflict or Harmony? An ITC Commissioner's Perspective," 56 *Antitrust Law Journal* 417 1987.

Ethier, Wilfred J. "Dumping," *Journal of Political Economy*, Vol. 90, No. 3, 1982.

Epstein, Barbara. "The Illusory Conflict Between Antidumping and Antitrust," 18 *Antitrust Bulletin* 1 1973.

Finger, Michael J. *Antidumping: How It Works and Who Gets Hurt*, University of Michigan Press, 1993.

Foer, Albert A. "The Goals of Antitrust: Thoughts on Consumer Welfare in the U.S.," Working Paper 05-09, American Antitrust Institute, 2005.

Foer, Albert A., and Robert H. Lande. "The Evolution of United States Antitrust Law: The Past, Present, and (Possible) Future," <http://www.antitrustinstitute.org/node/10194> 1999.

Garten, Jeffrey E. "New Challenges in the World Economy: The Antidumping Law and U.S. Trade Policy," *Congressional Record* Volume 140, Number 58 (Thursday, May 12, 1994).

Gifford, Daniel J. "Rethinking the Relationship between Antidumping and Antitrust Laws," *American University International Law Review* 6 No. 3, 1991.

Greenspan, Alan. "Antitrust." Paper given at the Antitrust Seminar of the National Association of Business Economists, 1962.

Hansen, Wendy L. and Thomas J. Prusa. "The Road Most Taken: The Rise of Title VII Protection." *The World Economy* Volume 18, Issue 2, March 1995.

Hillberry, Russell, and Philip McCalman. "What Triggers an Anti-Dumping Petition?"

- Finding the Devil in the Detail," Working Paper.
- Hoekman, Bernard M., and Petros C. Mavroidis. "Dumping, Antidumping and Antitrust," *Journal of World Trade*, 1996.
- Horsefield, J. Keith (editor). *The International Monetary Fund 1945-1965, Volume III Documents*, International Monetary Fund, 1969.
- Irwin, Douglas A. *Against the Tide: An Intellectual History of Free Trade*, Princeton University Press, 1997.
- Irwin, Douglas A. "The Rise of U.S. Antidumping Activity in Historical Perspective," IMF Working Paper, February 2005.
- Jackson, John Howard. *The World Trading System*, Massachusetts Institute of Technology, 1989.
- Kirkwood, John B. and Robert H. Lande. "The Fundamental Goal of Antitrust: Protecting Consumers, Not Increasing Efficiency." *Notre Dame Law Review*, Vol. 84:1 2008.
- Lande, Robert H. "The rise and (coming) fall of efficiency as the ruler of antitrust." *The Antitrust Bulletin*, Fall 1988.
- Lande, Robert H. "Beyond Chicago: Will Activist Antitrust Rise Again?" *The Antitrust Bulletin*, Spring 1994.
- Lynn, Barry C. *Cornered* John Wiley and Sons, 2010.
- Lynn, Barry C., and Phillip Longman. "Who Broke America's Job Machine?" *The Washington Monthly*, 2010.
- Mankiw, N. Gregory, and Philip L. Swagel. "Antidumping: The Third Rail of Trade Policy." *Foreign Affairs* Volume 84 No.4 July/August 2005.
- Markovich, Steven J. "U.S. Antitrust Policy." Council on Foreign Relations. Mar 2013. <http://www.cfr.org/business-and-foreign-policy/us-antitrust-policy/p29984>
- Mastel, Greg. *Antidumping Laws and the U.S. Economy*. M.E. Sharpe, 1998.
- Matsushita v. Zenith Radio Corp.* - 475 U.S. 574 (1986) :: Justia US Supreme Court Center. <http://supreme.justia.com/cases/federal/us/475/574/case.html#598> downloaded 3/6/2014.
- McGee, Robert W. "Antidumping Laws as Protectionist Trade Barriers: The Case for Repeal." *Policy Analysis* (Dumont Institute for Public Policy Research) No. 14 July, 1996.
- McGee, Robert W. "Some Ethical Aspects of Antidumping Laws" *Reason Papers* 25, 2000.
- Mitchell, Stacy. *Big Box Swindle*. Beacon Press, 2006.
- Moore, Michael A. "U.S. "facts available" antidumping decisions: An empirical analysis." *European Journal of Political Economy*, Vol. 22, No. 3, pp. 639-652, September 2006.
- Morgan, Clarisse. "Competition Policy and Anti-Dumping Is it Time for a Reality Check?" 30 *Journal of World Trade* 61 1996.
- Orbach, Barak Y. "The Antitrust Consumer Welfare Paradox." *Journal of Competition Law & Economics*, 7(1), 133-164, 2010.
- Page, William H. "The Ideological Origins and Evolution of U.S. Antitrust Law" in *Issues in Competition Law and Policy* (2008).
http://www.pepperlaw.com/publications_article.aspx?ArticleKey=318 downloaded December 16, 2013.
- Pitofsky, Robert. "The Political Content of Antitrust," 127 *University of Pennsylvania Law Review*, 4 April 1979.
- Priest, George L. "The Abiding Influence of The Antitrust Paradox: An Essay in Honor of Robert H. Bork." 31 *Harvard Journal of Law & Public Policy* 456 2008

Prusa, Thomas J. 1996. "The Trade Effects of U.S. Antidumping Actions." NBER Working Paper 5440, January 1996.

Prusa, Thomas J. and Susan Skeath. "The Economic and Strategic Motives for Antidumping Filings." NBER Working Paper 8424, August 2001.

Ruhl, Kim J. "Antidumping in the Aggregate." New York University Stern School of Business, March 2013.

Shin, Hyun Ja, "Possible Instances of Predatory Pricing in Recent U.S. Antidumping Cases," in *Brookings Trade Forum*, 1998.

Stewart, Terence P. "Why Anti-Dumping Laws Need Not Be Cloned After Competition Laws Nor Replaced By Such Laws," Centre for Applied Studies in International Negotiations Anti-Dumping and Competition Policy. Geneva, Switzerland, 11-12 July 1996

Stiglitz, Joseph E. "Dumping on Free Trade: The U.S. import Trade Laws," *Southern Economic Journal*, 64(2), 1997.

Sykes, Alan S. "Antidumping and Antitrust: What Problems Does Each Address?" in *Brookings Trade Forum*, 1998.

U.S. International Trade Commission, Office of Investigations. "Historical Case Statistics (FY 1980-2008)," February 2010.

Victor, A. Paul. "Antidumping and Antitrust: Can the Inconsistencies Be Resolved?" 15 *N.Y.U. Journal of International Law & Politics*. 339 1982-1983.

Viner, Jacob. *Antidumping: A Problem in International Trade*. Chicago: University of Chicago Press, 1923.

Wallace, Henry A. "The Danger of American Fascism." <http://newdeal.feri.org/wallace/haw23.htm> , downloaded December 17, 2013.

Wares, William A. *The Theory of Dumping and American commercial policy*. D.C. Heath and Company, 1977.

Willig, Robert. "Economic Effects of Antidumping Policy," in *Brookings Trade Forum*, 1998.

Wood, Diane P. "'Unfair' Trade Injury: A Competition-Based Approach," 41 *Stanford Law Review* 1153 1988-1989.

Wu, Mark. "Antidumping in Asia's Emerging Giants," *Harvard International Law Journal* Vol. 53, Winter 2012.

United States Permanent Mission to the World Trade Organization, "Observations on the Distinctions between Competition Laws and Antidumping Rules," *Working Group on the Interaction between Trade and Competition Policy*, WT/WGTCP/W/88, 28 August 1998.

Vandenbussche, Hylke and Maurizio Zanardi. "What explains the proliferation of antidumping laws?" *Economic Policy* January 2008.

Zanardi, Maurizio. "Antidumping, a Problem in International Trade," CentER Discussion Paper Series No. 2005-85, 2005.

Author contact: jbnere@verizon.net

SUGGESTED CITATION:

John B. Benedetto, "Placing economists' analyses of antidumping in an antitrust context", *real-world economics review*, issue no. 68, 21 August 2014, pp. 119-146, <http://www.paecon.net/PAEReview/issue68/Benedetto68.pdf>

You may post and read comments on this paper at <http://rwer.wordpress.com/comments-on-rwer-issue-no-68/>

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
Frederic S. Lee, 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 Radford, USA, The Radford Free Press
David Ruccio, USA, Notre Dame University
Immanuel Wallerstein, USA, Yale University

ISSN 1755-9472

EDITOR: Edward Fullbrook,

ASSOCIATE EDITOR: Jamie Morgan

PAST CONTRIBUTORS: James Galbraith, Frank Ackerman, André Orléan, Hugh Stretton, Jacques Sapir, Edward Fullbrook, Gilles Raveaud, Deirdre McCloskey, Tony Lawson, Geoff Harcourt, Joseph Halevi, Sheila C. Dow, Kurt Jacobsen, The Cambridge 27, Paul Ormerod, Steve Keen, Grazia Ietto-Gillies, Emmanuelle Benicourt, Le Movement Autisme-Economie, Geoffrey Hodgson, Ben Fine, Michael A. Bernstein, Julie A. Nelson, Jeff Gates, Anne Mayhew, Bruce Edmonds, Jason Potts, John Nightingale, Alan Shipman, Peter E. Earl, Marc Lavoie, Jean Gadrey, Peter Söderbaum, Bernard Guerrien, Susan Feiner, Warren J. Samuels, Katalin Martinás, George M. Frankfurter, Elton G. McGoun, Yanis Varoufakis, Alex Millmow, Bruce J. Caldwell, Poul Thøris Madsen, Helge Peukert, Dietmar Lindenberger, Reiner Kümmler, Jane King, Peter Dorman, K.M.P. Williams, Frank Rotering, Ha-Joon Chang, Claude Mouchot, Robert E. Lane, James G. Devine, Richard Wolff, Jamie Morgan, Robert Heilbroner, William Milberg, Stephen T. Ziliak, Steve Fleetwood, Tony Aspromourgos, Yves Gingras, Ingrid Robeyns, Robert Scott Gassler, Grischa Periono, Esther-Mirjam Sent, Ana Maria Bianchi, Steve Cohn, Peter Wynarczyk, Daniel Gay, Asatar Bair, Nathaniel Chamberland, James Bondio, Jared Ferrie, Goutam U. Jois, Charles K. Wilber, Robert Costanza, Saski Sivramkrishna, Jorge Buzaglo, Jim Stanford, Matthew McCartney, Herman E. Daly, Kyle Siler, Kepa M. Ormazabal, Antonio Garrido, Robert Locke, J. E. King, Paul Davidson, Juan Pablo Pardo-Guerra, Kevin Quinn, Trond Andresen, Shaun Hargreaves Heap, Lewis L. Smith, Gautam Mukerjee, Ian Fletcher, Rajni Bakshi, M. Ben-Yami, Deborah Campbell, Irene van Staveren, Neva Goodwin, Thomas Weisskopf, Mehrdad Vahabi, Erik S. Reinert, Jeroen Van Bouwel, Bruce R. McFarling, Pia Malaney, Andrew Spielman, Jeffery Sachs, Julian Edney, Frederic S. Lee, Paul Downward, Andrew Mearman, Dean Baker, Tom Green, David Ellerman, Wolfgang Drechsler, Clay Shirky, Bjørn-Ivar Davidsen, Robert F. Garnett, Jr., François Eymard-Duverney, Olivier Favereau, Robert Salais, Laurent Thévenot, Mohamed Aslam Haneef, Kurt Rothschild, Jomo K. S., Gustavo Marqués, David F. Ruccio, John Barry, William Kaye-Blake, Michael Ash, Donald Gillies, Kevin P. Gallagher, Lyuba Zarsky, Michel Bauwens, Bruce Cumings, Concetta Balestra, Frank Fagan, Christian Arnsperger, Stanley Alcorn, Ben Solarz, Sanford Jacoby, Kari Polanyi, P. Sainath, Margaret Legum, Juan Carlos Moreno-Brid, Igor Pauno, Ron Morrison, John Schmitt, Ben Zipperer, John B. Davis, Alan Freeman, Andrew Kliman, Philip Ball, Alan Goodacre, Robert McMaster, David A. Bainbridge, Richard Parker, Tim Costello, Brendan Smith, Jeremy Brecher, Peter T. Manicas, Arjo Klamer, Donald MacKenzie, Max Wright, Joseph E. Stiglitz, George Irvin, Frédéric Lordon, James Angresano, Robert Pollin, Heidi Garrett-Peltier, Dani Rodrik, Marcellus Andrews, Riccardo Baldissone, Ted Trainer, Kenneth J. Arrow, Brian Snowdon, Helen Johns, Fanny Coulomb, J. Paul Dunne, Jayati Ghosh, L. A. Duhs, Paul Shaffer, Donald W. Braben, Roland Fox, Stevan Harnad, Marco Gillies, Joshua C. Hall, Robert A. Lawson, Will Luther, JP Bouchaud, Claude Hillinger, George Soros, David George, Alan Wolfe, Thomas I. Palley, Sean Mallin, Clive Dilnot, Dan Turton, Korkut Ertürk, Gökcer Özgür, Geoff Tily, Jonathan M. Harris, Thomas I. Palley, Jan Kregel, Peter Gowan, David Colander, Hans Foellmer, Armin Haas, Alan Kirman, Katarina Juselius, Brigitte Sloth, Thomas Lux, Luigi Sapaventa, Gunnar Tómasson, Anatole Kaletsky, Robert R. Locke, Bill Lucarelli, L. Randall Wray, Mark Weisbrot, Walden Bello, Marvin Brown, Deniz Kellecioglu, Esteban Pérez Caldentey, Matías Vernengo, Thodoris Koutsobinas, David A. Westbrook, Peter Radford, Paul A. David, Richard Smith, Russell Standish, Yeva Nersisyan, Elizabeth Stanton, Jonathan Kirshner, Thomas Wells, Bruce Elmslie, Steve Marglin, Adam Kessler, John Duffield, Mary Mellor, Merijn Knibbe, Michael Hudson, Lars

Pålsson Syll, Korkut Erturk, Jane D'Arista, Richard Smith, Ali Kadri, Egmont Kakarot-Handtke, Ozgur Gun, George DeMartino, Robert H. Wade, Silla Sigurgeirsdottir, Victor A. Beker, Pavlina R. Tcherneva, Dietmar Peetz, Heribert Genreith, Mazhar Siraj, Ted Trainer, Hazel Henderson, Nicolas Bouleau, Geoff Davies, D.A. Hollanders, Richard C. Koo, Jorge Rojas, Marshall Auerback, Bill Lucarelli, Robert McMaster, Fernando Garcia-Quero, Fernando López Castellano, Robin Pope and Reinhard Selten, Patrick Spread, Wilson Sy, Esteban Pérez-Caldentey, Trond Andresen, Fred Moseley, Claude Hillinger, Shimshon Bichler, Johnathan Nitzan, Nikolaos Karagiannis, Alexander G. Kondeas, Roy H. Grieve, Samuel Alexander, Asad Zaman, L. Frederick Zaman, Avner Offer, Jack Reardon, Yinan Tang, Wolfram Elsner, Torsten Heinrich, Ping Chen, Stuart Birks, Dimitrios Koumparoulis, Arne Heise, Mark Jablonowski, Carlos Guerrero de Lizardi, Norbert Häring, William White, Jonathan Barzilai, David Rosnick, Alan Taylor Harvey, David Hemenway, Ann Pettifor, Dirk Helbing, Alan Kirman, Douglas Grote, Brett Fiebiger, Thomas Colignatus, M. Shahid Alam, Bryant Chen, Judea Pearl, John Pullen, Tom Mayer, Thomas Oechsle, Emmanuel Saez, Joseph Noko, Joseph Huber, Hubert Buch-Hansen, Brendan Sheehan, C P Chandrasekhar, Heikki Patomäki, Romar Correa, Piet-Hein van Eeghen, Max Koch, John Robinson, Oscar Ugarteche, Taddese Mezgebo, Donald Katzner

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=_NDUzNDkzMhwzMzIzODZ8cGFIX25ld3NAYnRpbmRlcm5ldC5jb218MTA5OTU4