sanity, humanity and science probably the world's most read economics journal

real-world economics review

Please click here to support this journal and the WEA

- Subscribers: **26,498** **subscribe** RWER Blog ISSN 1755-9472
- A journal of the World Economics Association (WEA) 14,588 members, **join**
- Sister open access journals: *Economic Thought* and *World Economic Review*

back issues

Issue no. 77

10 December 2016

Producing the RWER and running the WEA takes money.
By paying a WEA membership fee here, or by making a contribution here, you will be making them possible.

real-world economics review, issue no. 77
subscribe for free

Human growth and avoiding European disintegration: lessons from Polanyi[1]

Jorge Buzaglo [Independent economist, Stockholm]

Abstract
Many economists, including mainstream economists, have declared the necessity of an ambitious public investment program for Europe. The continuation of the present laissez-faire and austerity approach, they say, will deepen the dissatisfaction of European peoples with the European project. In effect, in the absence of inspiring progressive alternatives, there is the real prospect of nationalistic reactions everywhere, with fragmentation and the end of the EU (or even worse) as a consequence. At the bottom of EU-fatigue there is dissatisfaction with the present economic paradigm, which implies social and economic strain and stagnation for a large majority, and growing wealth and privilege for a tiny minority (the "one percent"). A new social and economic paradigm should be necessary to reverse the negative and dangerous trends that the now more than 30 years old neoliberal paradigm has implied. To be effective, a new social and economic paradigm should underlie the new, ambitious European investment program. The guiding general idea of such a new, hopeful paradigm and program should be the idea of *human growth*. Human growth can be understood as the growth of human capabilities — the all-round development of individuals and societies. The paper summarizes a dynamic simulation model for the efficient allocation of investment. Investment efficiency is understood in the sense of the intersectoral allocation of investments being able to fulfil desired patterns of capability growth. That is, efficient investment programs are those that maximize the capacity of the production system — and in general, the socioeconomic system — of generating human growth.

Introduction

In this short paper I set out some lessons and insights taken from Polanyi's *The Great Transformation*. I do so in order to make the case for *human growth* as a way to orient public investment in Europe and beyond. This is one way to think about a transformed Europe in which economies are re-embedded within social and political systems. I outline a simple, consistent investment planning model for human growth in Europe. The model is a *policy simulation model*. That is, it does not pretend to explain or replicate the almost infinite complexity of a real modern economy evolving in time, but to describe the plausible mechanisms governing the restricted parts of the economy that are relevant for efficient investment policy planning. The public sector has an important role in determining the direction and rate of structural change and growth in the economy. The role of public investment policy can be to foment structural transformation and human growth.

[1] New version of a paper presented at The 22nd Workshop on Alternative Economic Policy in Europe, on *The European Union: The Threat of Disintegration*, organized by the EuroMemo Group, and jointly hosted with the Faculty of Economics at the University of Coimbra (September 2016). I would like to thank the organizers, and participants at the workshop on Macroeconomic and developmental policies to challenge austerity and uneven development. I would also like to thank comments from Stefan de Vylder. The present version benefited greatly from the changes suggested by Jamie Morgan.

1. The concept of human growth

The idea of human growth self-evidently includes the term *growth*. This is a loaded term. Growth has positive associations in the presently prevailing ideological context of neoliberalism. However, its link to human flourishing is questionable. In *human growth,* growth is qualified: human growth is not any kind of growth. Human growth is the kind of growth that increases the universe of life possibilities of all persons in society (it is individual yet relational flourishing). Human growth is also the kind of idea, which is apt to evoke different associations and aspirations in different people, and may open the space of intellectual discovery and debate away from the negative and sterile prevailing view of human beings as selfish and narrow-minded *homo oeconomicus.*

Of course, the debate on human growth has been long conducted in economics. The dominant focus has been the limitations of — and alternatives to — GDP as an indicator of society's welfare, or of any other meaningful notion of socioeconomic advancement. The discussion about the limitations of, and alternatives to, GDP is in fact a discussion about what really counts when the advancement of a society is to be measured with some quantitative accuracy. Also, besides this debate referring to social accounting and the economics of welfare, the wide literature on the idea of *human development* has referred to the same problematic, from a more conceptual and philosophic point of view.

In our sense, the most fruitful notion of overall socioeconomic advancement is the idea of human growth. I define human growth as the expansion of every relevant domain of exertion of human capabilities (faculties, aptitudes, abilities) for every individual in society.

The idea of human growth as the enlargement of human capabilities is also not new. In different forms and to different degrees, the idea has been present in most philosophical and ethical traditions all over the globe. In the West in particular, the paramount ethical and political goal of human growth as capability expansion has been postulated within both the liberal and the socialist social-philosophical frame of ideas (for the liberal, see e.g. Nussbaum 2011, Sen 1999; on the socialist, e.g. Burkett 2005, Buzaglo 2014).

At the international level, a highly relevant illustration of the general recognition of human growth and capability goals, is the recent adoption by the United Nations of a set of 17 Sustainable Development Goals (SDGs), with 169 specific targets to be fulfilled by all countries through to 2030 (United Nations 2015). These goals and targets represent, directly or indirectly, a commitment for every country to pursue a specific set of human capabilities.

2. Lessons from Polanyi

> "Under the gold standard the leaders of the financial market are entrusted, in the nature of things, with the safeguarding of stable exchanges and sound internal credit on which government finance largely depends. The banking organization is thus in the position to obstruct any domestic move in the economic sphere which it happens to dislike, whether its reasons are good or bad. In terms of politics, on currency and credit, governments must take the advice of the bankers, who alone can know whether any financial measure would or would not endanger the capital market and the exchanges **…** The financial market governs by panic" (Polanyi, 1944, p. 229).

The present globalization wave is a financialization wave. The previous waves described in Karl Polanyi's *The Great Transformation* also involved financialization. The financial sector appears as the major and most consistent leader and beneficiary of the general marketization policies of deregulation, liberalization and privatization. One of the main lessons of *The Great Transformation* is that previous globalization waves radically undermined established systems of social relations, thus provoking, in time, socially and/or nationally "protectionist" reactions, within a "double movement" (for more on Polanyi, see Zaman, 2016).

As Polanyi also observes, political reactions could be extreme:

> "After 1930 market economy was in a general crisis. Within a few years fascism was a world power" (Polanyi 1944, p. 242).

It seems that, if the lesson from *The Great Transformation* holds also for the present globalization wave, we should expect social and national reactions trying to "re-embed" the de-regulated, unbridled forces of the market within new networks and rules of social life. Nationalism seems to be the more primitive and irrational form of "protectionist" reaction. An eroded and divided society, in which political parties and elites have lost legitimacy, is unified around the myth of the great — and abused — Nation. Minorities become potentially treasonous; other nations become potentially enemies. Fascism can present itself in several different guises, from the more folkloristic and semi-democratic, to the aggressively imperialistic and genocidal. We are perhaps already witnessing different forms of implementation of evolving, "post-modern" fascism.

The social protectionist, or socialist reaction, on the other hand, is a more rational reaction, which is based on an analysis — by the social movements and their intellectuals — of the nature of the existing social conflict and its possible progressive resolutions. The socialist response is therefore less visceral and more difficult. It requires generous class solidarity, even at the international level. It requires an ability to form wide programmatic alliances, and to unify the plurality of social forces characteristic of modern, late capitalism. It requires endurance and patience, clarity about long term interests and objectives, and a capacity to postpone and negotiate for further advancement.

3. Human growth as a new European paradigm

The evolution of Europe after 2008 does not encourage optimism about the future of the "social-protectionist" response to the crisis. In spite of expectations opened directly after the crisis, neoliberalism and austerity politics are entrenched. Perhaps even more seriously, there are few widely accepted and clearly articulated social projects to propose an alternative to crude, neoliberal capitalism.[2]

It seems necessary, therefore, to provide arguments and alternatives for a renewed socialist project in Europe that can appeal to the social consciousness of Europeans. Such a social movement must be able to articulate a common, hopeful strategy for European transformation. A vision for change in Europe should imply a new economic and social

[2] However, there is the attempt by the Democracy in Europe Movement 2025 (DiEM25) movement to develop a Progressive Agenda for Europe, which will include proposals for a European New Deal and a European Constitution (see diem25.org).

paradigm beyond neoliberalism, including democratic reform of European institutions. A focus on Human Growth could be such a new, inclusive paradigm for Europe.

Given the present EU-fatigue and distrust, a program for institutional reform and democratization should perhaps consider a first phase of increased autonomy for national parliaments and governments. However, the crucial devolution of autonomy and sovereignty would only come with the transfer of the key aspects of economic policy — such as fiscal, and in particular, investment policy — from the financial markets to national and EU programs conceived in terms of human growth.

According to Polanyi:

> [T]he victory of fascism was made practically unavoidable by the liberals' obstruction of any reform involving planning, regulation, or control. (Polanyi 1944, p. 257)

The direction and content of socioeconomic development in the EU have been essentially determined by market laissez-faire, ultimately dictated by the financial markets' irrational, manic-depressive psychology. The institutions and the economics of the orthodoxy, which form the constitutional ideological ground of the EU inescapably lead to mass unemployment, large and increasing inequalities of incomes and wealth. This in turn leads to increasing alienation of large sections of the population, and finally to the kind of nationalistic reaction that we see today.

To avoid a protracted crisis and the trend toward disintegration and internecine conflict, institutions and policies of the EU need to be transformed. Most urgently, a massive investment plan should clearly signal the end of the current orthodoxy and austerity politics. This would indicate a new orientation of the EU, away from neoliberalism, and towards a more representative and deliberative form of accountable democracy consistent with a new paradigm of human growth. Human growth could be the idea that reunites the European nations behind the EU project, and replaces the neoliberal — in fact, anti-democratic — paradigm.

Essential for progressive change is, therefore, to provide a powerful set of ideas, capable of articulating the risks involved in present trends and the potentials of a transformed Europe. Moreover, a well-defined and well-specified program for paradigm change in Europe could be important not only for Europe, but also to induce similar changes at the global level, thereby reducing global economic instability and international political/security conflict and tension.

4. "Big push" investment planning for human growth

> Since the introduction of "functional finance" in all important states, the directing of investments and the regulation of the rate of saving have become government tasks. (Polanyi 1944, p. 252)

Clearly, a "big push" investment program focussed on human growth in Europe would be consistent with the rationale of the SDGs, which indicate a necessity to transcend a neoliberal phase of destabilizing and haphazard directing of investment by the financial markets. For example, neoliberalism seems manifestly incapable of delivering environmental sustainability

(a core issue for the new SDGs). Moreover, the current general context remains one of stagnation, unemployment, growing income and wealth disparities, and increasing social and political tensions, national and international.

A European investment program focussed around human growth need not be limited to the goals of the UN 2030 Agenda. It could include a more ambitious and wider set of goals and targets, specifically inspired by the European problematic, and attacking the social and economic causes behind growing nationalism, xenophobia and racism (mass unemployment, growing social exclusion and alienation, regional imbalances and disparities, etc.).

What we are speaking about is *investment planning.* This is a subject that is anathema to the official economics of the neoliberal age. How society can and should rationally and effectively allocate investment among different ends through public processes has been a forbidden subject in mainstream economics, because it is assumed that in a market society investment allocation is best done by private investors in the market. It is even "demonstrated" that it is impossible to form a social choice function capable of representing a society's preferences. That is, it is impossible to construct a function able to guide economic policy and planning choices. According to mainstream theory, any such attempt should be flawed in some important sense (based on Arrow's famous "impossibility theorem"). However, for many decades now, most economies have had large public sectors, and have collectively (often democratically) allocated an important share of total investment through the public sector. This is to do something that for economics is apparently impossible, or at least seriously flawed.[3] The kind of collective decision-making that concern us cannot be done by a social choice algorithm, constructed by some clever mathematician. It must be done through a transparent and democratically accountable deliberative political process. Models and algorithms, at best, may help to shed some light on the character of the problem to be solved by the decision process.

5. An investment policy simulation model

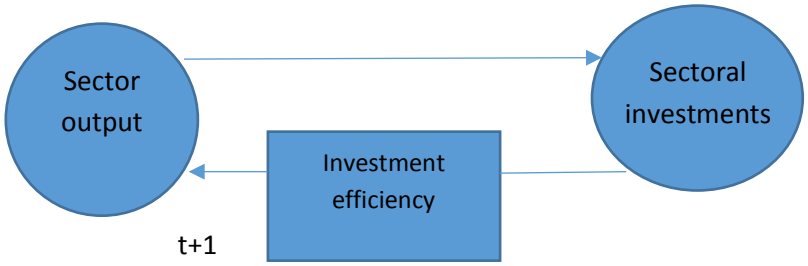
Let us now try and present an outline of a simple, consistent investment-planning model for human growth in Europe. The model is a *policy simulation model*, that is, it does not pretend to explain or replicate the almost infinite complexity of a real modern economy evolving in time, but to describe the plausible mechanisms governing the restricted parts of the economy that are relevant for efficient investment policy planning. In particular, we are interested in the relationship between means and ends, or instruments and objectives. That is, between sectoral investments and human capabilities. The model is largely based on dynamic input-output and social accounting matrix theory (for a more detailed description, see Buzaglo 2016).

The basic causal structure and law of motion of the model is given by the classical theory of growth or capital accumulation. In every sector of activity, growth depends on investment, for

[3] Was the (pseudo-)Nobel Prize given to Kenneth Arrow a flawed collective decision? Seriously, not only the Nobel Committee make collective decisions: most private investment decisions are also taken by collective bodies such as corporations. Even *individual* investment decisions are (flawed?) collective decisions, taken by about 86 billion individual neurons, in turn composed of billions of molecules, in turn composed of many atoms, in turn composed of a large and growing number of particle-waves … and so on, with no end in view. Science could paraphrase Madam Thatcher: "There is no such a thing as individuals."

given factors of output response to investment, or (sectoral) investment efficiency parameters. As a dynamic input-output model, the model follows a (disaggregated) "Harrod-Domar" law. Figure 1 shows an intuitive visualization of the basic law of motion over time.

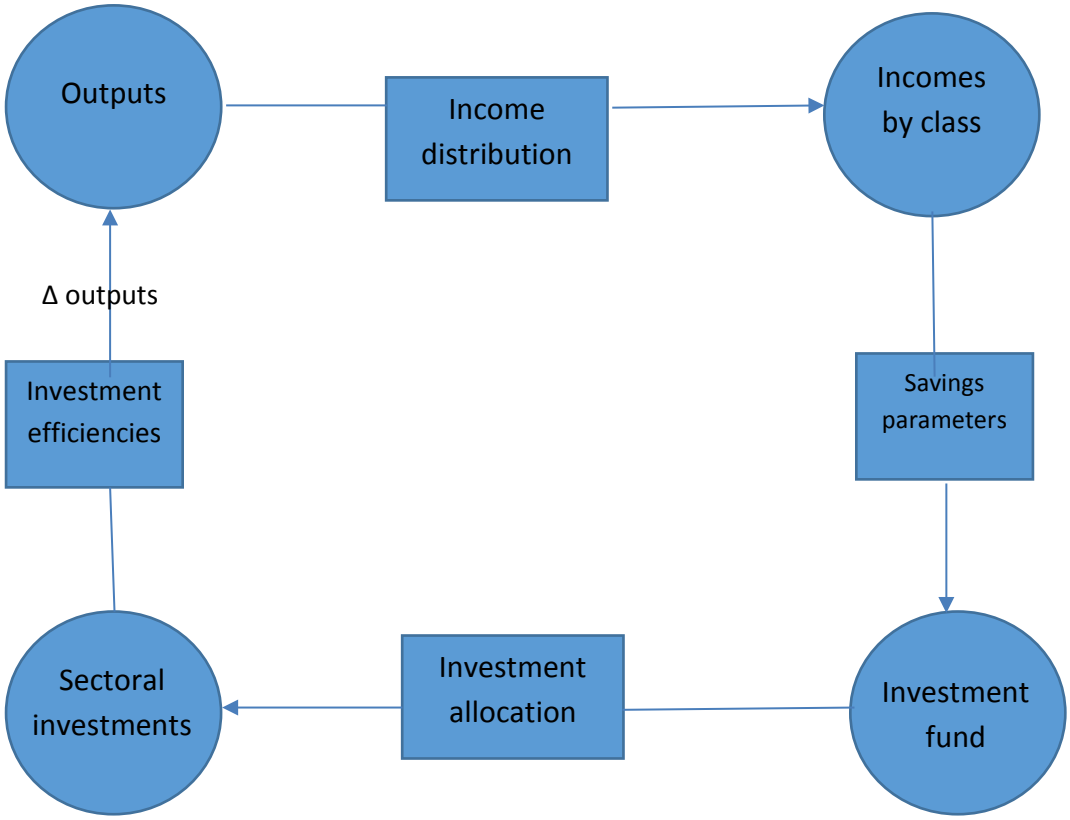
Figure 1



Growth thus depends on the level and distribution of investment. Economic policy, in particular fiscal policy, influences the level and distribution of investment, thus influencing sectoral rates of growth — the level and composition of output over time. Knowingly or unknowingly, willingly or unwillingly, the public sector has an important role in determining the direction and rate of structural change and growth in the economy. The role of public investment policy can be to induce status quo (or even involutionary) growth, or to foment structural transformation and human growth.

In order to visualize both sides of fiscal policy, i.e. fiscal income and expenditure policies, it is convenient to also represent income generation and distribution. Figure 2 shows how incomes generated in production are distributed among different income groups or classes (and the government). The public sector redistributes incomes through taxes and subsidies, and finances its expenditures, including investments.

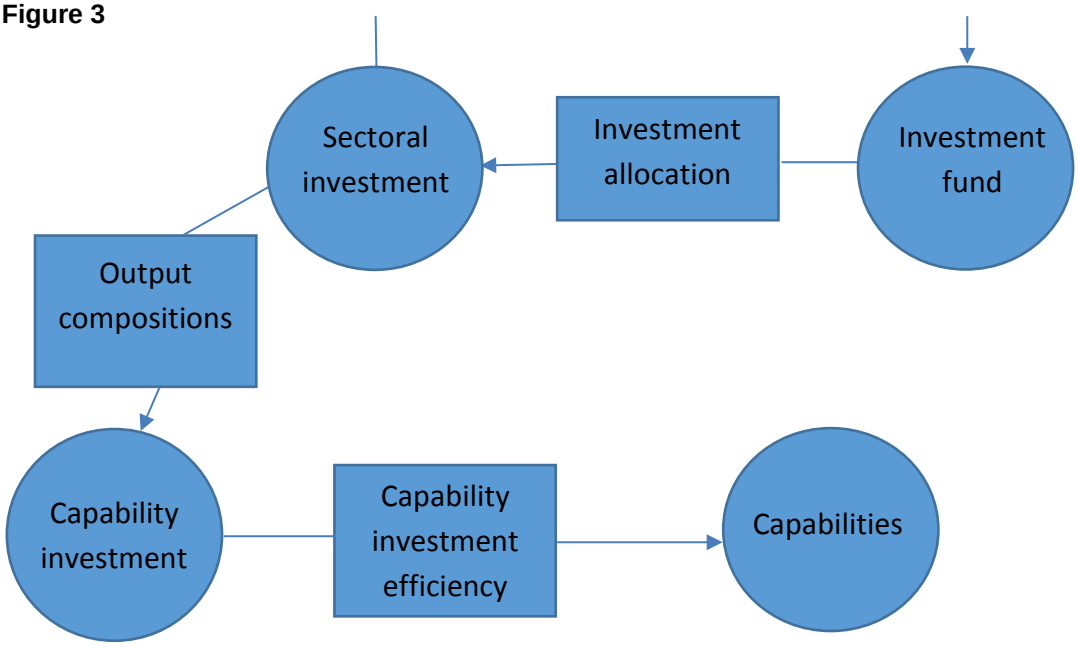
Figure 2



Both income distribution and investment allocation — the upper and lower boxes — can be understood as policy functions: income distribution and investment policies may both be used to achieve desired patterns of human growth. However, in this presentation, devoted to the urgent necessity of a "big push for human growth" in Europe, we focus on the planning of investments with that aim.

In order to be able to plan the allocation of investments for human growth, we need a link between the production system shown in Figure 2 with the space of human capabilities. There exists a link between investments in the producing sectors and investments in capability-creating activities. We want to know how investment in goods and services in the producing sectors should be re-oriented in order to achieve desired human growth goals. In order to do that, we need to know in what proportions investments in the different capability-creating activities are composed of goods and services from the different producing sectors. And/or also, we need to know in what proportions investments in the different producing sectors contribute to the formation of each one of the different categories of capabilities.

Figure 3



In Figure 3 we show the connection between investment in the output producing sectors with investment and growth in human capabilities. The upper level of the figure shows the lower part of Figure 2, describing how investment policy orients the allocation of investment among the producing sectors, thus modifying output structure over time. In the lower levels of the figure these investments, in turn, are transformed in capability creating investments, according to known proportions of the output compositions of capability investments.

The last step is to determine the effect of the investments in the different capability creating activities in the growth of their respective human capability index. Given known investment/capability parameters, or capability investment efficiency parameters, the (effect of capability investments on the) growth of the different capabilities is determined.

The message of Figure 3 is then that the level and distribution of investments in the producing sectors influences the growth and composition of capabilities. Also, that different goals of human capability growth can be approached by steering investments in the producing sectors so that the goals are fulfilled to such extent as is possible. Investments conceived in this manner could be called "human-growth-efficient," or simply "human-efficient" investments.

Conclusion

We have taken from Polanyi's *The Great Transformation* some lessons and insights, which were once shared by many people, but have almost disappeared from the intellectual horizon (at least in economics). The consequence has been more than thirty years of universal marketization and commodification. A first lesson is that globalization waves undermine and unravel established forms of economic life, leading eventually to social and/or national reactions. The nationalistic reaction, in particular, may take gravely deleterious forms, such as fascism. Another important lesson is that entrenched [neo-] liberalism in policy, and the political prevalence of financial interests in society, can block any possibility of regulation in the general interest, so opening the way for chauvinism and fascism.

This paper articulates a socialist response to the crisis. The common theoretical and ideological ground for a wide program for progressive change in Europe should be the idea of *human growth*. A key component of this program for change should be a large, "big push" investment program focussed on human growth. Adoption of such a program may signal the end of the neoliberal age and the start of a new, more hopeful paradigm in Europe.

The investment policy simulation model in the paper shows the possible logical structure of programming investment for human capability growth. It is possible to reorient investment — from the stagnation and inequality induced by the financial markets, to economic regulation aimed at human growth. As Polanyi also highlights, economic regulation and investment planning were commonplace before the neoliberal age. That is, during the "Golden Age of Growth." To *summarize* our *main policy points*:

1) To counter growing nationalist and reactionary trends, it is crucial to articulate a common, hopeful social alternative to rampant neoliberal capitalism. The alternative should be a new, broad vision for change in Europe — a new whole economic and social paradigm beyond neoliberalism. That new paradigm beyond neoliberalism should be constructed around the idea of human growth: *a new paradigm for human growth in Europe.*
2) Human growth can be understood as the expansion of every relevant domain of exertion of human capabilities (faculties, aptitudes, abilities) for every individual in society — the all-round development of individuals and societies. The extant Sustainable Development Goals (United Nations 2015) constitute an operational and uncontroversial set of 17 capability goals to be fulfilled by all countries through to 2030. These goals should be enlarged, and adapted to the specific European problematic.
3) A massive, "big push" investment program for human growth in Europe should signal the end of the neoliberal age and the beginning of the new, hopeful paradigm of human growth.
4) The investment policy planning model of the paper shows how investment in the different sectors of the economy can be (re)oriented over time in order to fulfil

capability targets — that is, how an investment program for human growth could be conceived, and its fulfilment followed in time.

References

Burkett, P. (2005) "Marx's Vision of Sustainable Human Development," *Monthly Review*, Volume 57, Issue 05 (October). (http://monthlyreview.org/2005/10/01/marxs-vision-of-sustainable-human-development/)

Buzaglo, J. (2014) "The ℵ Capability Matrix: Some observations on GDP and the economics of human development," *Real-World Economics Review*, issue no. 66. (http://www.paecon.net/PAEReview/issue66/Buzaglo66.pd.)

Buzaglo, J. (2016) 'Expanding Human Capabilities: Lange's "Observations" Updated for the 21st Century.' *Economic Thought*, 5.2, pp. 1-11. http://www.worldeconomicsassociation.org/files/journals/economicthought/WEA-ET-5-2-Buzaglo.pdf

Nussbaum, M. (2011) *Creating capabilities: the human development approach.* Harvard University Press, Cambridge, Mass.

Polanyi, K. (1968 [1944]) *The Great Transformation.* Beacon Press, Boston.

Sen, A. (1999) *Development as Freedom*, Oxford University Press, Oxford.

United Nations (2015) Transforming our world: the 2030 Agenda for Sustainable Development. (https://sustainabledevelopment.un.org/post2015/transformingourworld)

Zaman, A. (2016) "The methodology of Polanyi's Great Transformation," *Economic Thought* 5.1, pp. 44-63

Author contact: j.buzaglo@telia.com

SUGGESTED CITATION:
Jorge Buzaglo, "Human growth and avoiding European disintegration: lessons from Polanyi", *real-world economics review*, issue no. 77, 10 December 2016, pp. 2-10,
http://www.paecon.net/PAEReview/issue77/Buzaglo77.pdf

real-world economics review, issue no. 77
subscribe for free

Uncertainty about uncertainty: the futility of benefit-cost analysis for climate change policy

Mariano Torras [Adelphi University, NY, USA]

Abstract
The moral argument for aggressively mitigating continued greenhouse emissions is very strong. The economic case is not as clear, however, owing to great uncertainty about the benefits and costs. Three distinct types of uncertainty – predictive, valuational, and moral – hamper and severely limit the usefulness of social benefit-cost analysis by introducing irreducible subjectivity into the analysis. The complex nature of climate change compels us to rethink the role of economics in policy and better balance it with qualitative reasoning and judgment.

Keywords climate change, uncertainty, subjectivity, social benefit-cost analysis, social discount rate, ethics

1. Introduction

Mounting evidence that anthropogenic climate change is a true phenomenon by now seems irrefutable. While climate and earth scientists have long been calling attention to indicators that climate change is caused by humans, economists have also recently grown more concerned. Yet in their case, the science behind climate change or its likelihood and extent has been of only secondary – albeit hardly peripheral – interest. The primary focus for economists has been on the potential social benefits and social costs of addressing climate change.

Many economists find climate change to be a minor problem when compared to other challenges like hunger or promoting global trade.[1] The belief largely stems from unrealistic assumptions about the modest impact on the global economy of disruptions in agricultural production, as well as about the likely economic consequences of even extreme temperature increases (see, e.g., Nordhaus and Boyer, 2000). But this is not all. The publication of Nicholas Stern's (2007) Review, which claims that the social benefit of aggressive climate change policy is about 100 times greater than that estimated by Nordhaus, helped spark a debate among economists. The social discount rate, to be discussed in detail later, is at the center of this disagreement.

Growing awareness and recognition of the problem of climate change has recently opened the door to at least the possibility of substantive progress in policy. The landmark multinational agreement at the December 2015 Paris Summit (hereafter the Paris Agreement) is an especially welcome sign after decades of prevarication. Yet it is merely a first step, and much of what happens in the future regarding climate change depends on the resolve of policymakers in making the necessary sacrifices to comply with the agreement. Morgan (2016) indeed questions whether there we have seen any fundamental change that could

[1] See e.g., Cline (2004), Nordhaus (2007), and Tol (2003, 2009). Cline and Tol are among a panel of experts (including a few Nobel laureates) that form the Copenhagen Consensus think tank. It is possibly the most notable purveyor of the notion that climate change is relatively unimportant compared to other global challenges.

translate the conditional success of the Paris Agreement to an actual success that might realize its goals.

The degree of commitment of national leaders will, in the long run, come down to their perception of the relevant benefits and costs. This is the principal concern in this paper. I argue that while the evidence in support of climate change is convincing, attempting to gain an unrealistically *precise* understanding of the benefits and costs of ameliorative policy for use in social benefit-cost analysis is potentially worse than useless. Even with a superior understanding than we presently have, we would still not know what dollar values to put on the relevant benefits and costs of climate change policy. And even if we surmounted this obstacle, there is no objective basis for a rate at which to discount benefits and costs into the future. Quantitative precision is likely, therefore, to lead us astray. The challenge, ultimately, will be about coming to recognize that ethics, politics, and other subjective considerations may trump crude economic analysis in determining the appropriate climate policy.

2. What do we mean by uncertainty?

Mitigation vs adaptation

Thirty to forty years ago, climate change (then referred to as 'the greenhouse effect' or 'global warming') was mostly considered a fringe topic. Skepticism prevailed, and those believing climate change to be a real phenomenon tended mostly to see it as a long-term problem. Even today, despite incontrovertible evidence supporting the reality and the imminence of climate change, uncertainty over the details has sustained purveyors of the notion that the idea is a hoax.

There is no shortage of detailed evidence that climate change is happening; it can be found in the numerous quantitative studies that have been published in recent years (Hansen and Sato, 2016; Pachauri and Reisinger, 2007). Yet given the massive economic sacrifice potentially implied by aggressive mitigation policy, the institutional pressure to obfuscate – referred to by Best (2012) as 'bureaucratic ambiguity' – is understandable.[2] Along similar lines, the idea that we cannot avoid either economic or environmental 'pain' in the future is an example of what Rayner (2012) calls 'uncomfortable knowledge' which, he argues, calls for a strategic mix of denial, dismissal, diversion, and displacement.[3]

While there may be no way of *unequivocally* linking massive floods, storms, and droughts in recent years to climate change, many nevertheless regard such events as sufficient circumstantial evidence in support of the climate change hypothesis. In statistical parlance, we might say that the climate change hypothesis has over the past 40 years switched from

[2] Obfuscation on matters of policy importance often – if paradoxically – take the form of quantitative (though often inaccurate) precision on a subset of relevant factors. Substantial uncertainty underlying climate change is concealed through the misleading use of precise figures. For more on this, see also Torras and Surie (2015).
[3] So, for example, the scandal over some ostensibly counterfactual emails suppressed by scientists supports efforts to deny the reality of climate change; undue emphasis on a recent retraction by the Intergovernmental Panel on Climate Change (IPCC) of its own overestimates of the rate of Himalayan glacial melt is an example of diversion. For detailed discussion on the subtle differences, see Rayner (2012).

'alternative' to 'null.' There is indeed no clearer sign of increasingly widespread acceptance of the hypothesis than the Paris Agreement of December 2015.[4]

The natural conclusion would appear to be that we should now be aggressively limiting further carbon emissions, at least until it could be demonstrated that doing so would not be worthwhile.[5] Yet it is one thing to accept science's interpretation of the evidence; it is quite another to claim to know things about future benefits and costs that we simply do not. For example, *how much* will climate change in terms of, say, average global temperature, and how long will it take to do so? What will be the effects on human societies? What can we expect to be the extent of the overall damage? Finally, and possibly most important, is catastrophic, even human-extinguishing, change possible – and if so, with what probability?

We have no unambiguous answers to these important questions. It is not even clear that we are, over time, approaching answers.[6] Climate change studies and data abound, but we can mostly only speculate on how they should be interpreted. Skeptics are actually correct in stating that *we do not know* if sacrifice today – in terms of reducing material and energy flows, slowing down the economy, etc. – would be 'worth it' in the long run. We might naturally assume that aggressive mitigation is vital and urgent (and, ethically speaking at least, such a claim would seem beyond reproach), but from an economic standpoint, it *could* make more sense to instead invest in adaptive technologies – dikes and seawalls, floating cities, etc. – aimed at ensuring survival.

Such a perspective confronts the uncomfortable possibility that substantial future damage is already irreversible. Would it not then be more economical and practical to spend money preparing for higher sea levels and hotter temperatures? While defense of a continued 'business as usual' strategy (hereafter BAU) – i.e., doing nothing now and adapting to climate change in due course – might mostly be politically motivated, unequivocal hard evidence to plausibly challenge it remains elusive.

The net benefit of climate change policy

Some researchers (e.g., Ackerman and Munitz, 2016; Hope, 2008) seek to reliably approximate the true social cost of carbon (SCC) as a basis for tax policy on continued carbon emissions. The problem with a tax is that it merely changes incentives but fails to strictly limit emissions. There is, therefore, no way of ensuring that any given emissions threshold is not breached as a consequence of an insufficiently large tax.

Because of this rather serious limitation, many prefer the alternative of a tradable carbon permits scheme. But there are also problems with a tradable permits system, such as fairness in the permit allocation or ensuring compliance with emissions limits. In theory, the tax on a unit of carbon emitted should equal the price of a permit to emit a unit, though in practice this

[4] Recognition of the problem is, sadly, one small step in addressing it. Ertürk and Whittle (2015) argue that rich and poor countries have different incentives, with the former wanting to export mitigation costs to the latter instead of bearing most of the responsibility. Spash (2016) notes the inconsistency between the Paris Agreement and national governments' commitment to economic growth at all costs.

[5] That emissions of methane, CFCs, and other greenhouse gases should also be severely curtailed hardly bears mentioning. Rather than repeating this point, the reader should consider my mention of carbon as meaning the carbon equivalent of the different greenhouse gases.

[6] This is increasingly recognized by establishment economists such as Nicholas Stern (2013), who argues that the dominant economic modeling approach is likely to seriously underestimate the downside risks associated with climate change.

is unlikely. Nevertheless, the debate over which is better is beyond our present scope. I am interested in the question of how we decide on the proper balance between mitigating and waiting (adapting), independent of the policy mechanisms chosen.

Economists frequently rely on social benefit-cost analysis to evaluate policy options, and here it is helpful to look at how the problem is often framed. Consider the following net benefit function,

$$NB = \sum_{t=1}^{\infty} \frac{B_t[\phi_t(\mu_t), \tau_t(\alpha_t)] - \mu_t - \alpha_t}{(1+d)^t} \qquad (1)$$

where NB stands for the discounted sum of social (or global, if you like) net benefits over time.

In this equation, μ_t represents mitigation cost and α_t adaptation cost, both in time *t*. The benefits from climate change policy (B_t) are a function of ϕ_t, which stands for the diminution in climate change compared to BAU, and τ_t, a measure of the extent to which we technologically adapt to climate changes (e.g., hotter temperatures, higher sea levels). ϕ_t and τ_t are, respectively, functions of μ_t and α_t.[7] Finally, *d* represents the social discount rate.

In theory, if the net benefit were positive – i.e., if Equation 1 had a value greater than zero – it would suggest that the specified combination of mitigation and adaptation policy is worthwhile. A negative value would imply the inverse, namely that we would have spent too much money and that BAU would have been better. The problem – and it is a big one – is that there are an *infinite number of possible* ϕ's, depending on the magnitude, balance, and timing of both mitigation and adaptation policies.

The three dimensions of uncertainty

It is widely recognized that the uncertainty regarding ϕ_t and its relationship to μ_t is enormous (egs., Allen and Frame, 2007; Rosen and Guenther, 2014; Woodward and Bishop, 1997). But there exist two other matters about which it is nearly impossible to be quantitatively precise. In addition to what I call *predictive* uncertainty, we confront *valuational* uncertainty – which is the problem of how to assign monetary values to the uncertain damages (or the uncertain benefits to averting damages) – and *moral* uncertainty – the quandary of how much relative importance to grant future generations of humans.

Only the first of the preceding three categories is generally treated as 'uncertainty' in the economics literature, and with good reason. Allowing too much uncertainty into economic analysis would weaken the claim that economics can be regarded as anything even resembling a precise science. Economists for the most part avoid even mentioning uncertainty; Hodgson (2011) notes that the frequency with which the word has appeared in articles in leading economics journals has been on a steady decline for decades.

Calling attention to valuational and moral uncertainty thus serves two purposes. First, doing so provides a more candid assessment of the limitations of orthodox economic analysis,

[7] While ϕ_t is shown as a function of μ_t, we should expect the two to mutually cause each other. I omitted the reverse link, as well as other likely relationships (including the effect of ϕ_t on α_t) to keep things reasonably simple.

especially when it comes to an issue as critical as climate change. While it has contributed a variety of widely applied concepts and models, there are strong grounds for questioning just how relevant these have been, especially in an environmental context. Controversy on this precise question has resulted in a split between environmental and ecological economics, and then also within ecological economics between those more or less critical of the methodologies that underpin many mainstream approaches.

Second, focusing on valuational and moral uncertainty lays bare the fact that subjectivity in analyzing climate change is inevitable, complicating matters much more than would be evident by focusing exclusively on the complexities of global climate. The more transparent we can be about this, the less temptation there will be to behave as if assigning arbitrary numbers will 'fix' the problem.

3. Predictive uncertainty: Risk, uncertainty, and ignorance

Climate, complexity, and unpredictability

Predictive uncertainty refers specifically to uncertainty about future events. It mostly relates to climate outcomes (ϕ_t), which are in part a consequence of mitigation expenses (μ_t). We do not know what climate change future ($\Phi = \int_{t=0}^{\infty} \phi_t$) awaits us, both because we have insufficient ecological and climatological awareness, and because we do not have adequate information on the response functions that determine μ_t, and α_t. Such uncertainty is not primarily about the market price of mitigation or adaptation technologies, but about the nature of the technologies themselves. Since we do not know what technologies will even be necessary in the future, we know less about how the climate will have responded to our policy action (or lack thereof) the further into the future we look. We cannot, moreover, predict how responsive our collective political will would be to unpredictable future events.

Future climate change outcomes and mitigation and adaptation policy costs are 'positive' issues in the sense that there are no value judgments involved. Despite enormous uncertainty, they are variables for which prediction entails pure, objective analysis. Unfortunately, the absence of value judgments does not make the task of accounting for predictive uncertainty noticeably easier.

Mechanistic models representing a phenomenon that is more or less deterministic often lend themselves nicely to predictive methods. With the aid of computers, today we can identify the relevant differential equations, plug in the known quantitative inputs, and arrive at precise conclusions regarding future outcomes. Yet with the enormous uncertainty and complexity that it entails, climate change is decidedly anything *but* mechanistic. Gowdy (1994), for example, has argued that an evolutionary modeling scheme better characterizes the global climate. While differential equations are usually also appropriate tools for such models, predicted outcomes are often highly sensitive to even minute changes in one or more of the relevant inputs. Uncertainty over details might therefore translate to even greater uncertainty over possible outcomes.

As noted by Weitzman (2009), there are two things we do not know: how much greenhouse gas will be emitted, and what the 'scaling factor' is. Uncertainty over the latter, which is

essentially the product of many variables, is far greater.[8] The complex interplay of different factors often gives rise to positive feedback loops. One such case is the melting of planetary ice sheets due to higher global temperatures. We know that since ice reflects solar energy, its widespread melting causes the earth to absorb more solar heat, causing further melting in what is known as ice-albedo feedback. But we do not know *how much* more, nor with what long-term effect. We also have good reason to suspect that increased temperatures will eventually melt sufficient permafrost to release sequestered methane deposits into the atmosphere, again reinforcing the temperature increase. But again, we have almost no idea the magnitude of the effect.

We could say the same about our knowledge of the role of tropical forests and deforestation; we understand the basic relationships but are utterly incapable of accurately forecasting long-term climate effects. Herein lies the fatal weakness of most climate models: while they are very useful and effective and providing insight and explaining phenomena at a basic level, they are unreliable, because of the immense complexity involved, for predicting or forecasting.

Uncertainty and ignorance

Different degrees of predictive uncertainty prevail, and clarifying the differences further illustrates the challenge presented by climate change. For example, a situation of 'weak' uncertainty (Table 1) commonly refers to cases where the outcome is not known *a priori* but all the possibilities and their respective probabilities are.[9] To be sure, for a problem as complex as climate change the assumption of weak uncertainty is wildly unrealistic. Yet in an era in which precise quantitative conclusions are at a premium, it becomes almost compulsory to *assume* weak uncertainty in order to facilitate – or even make possible – certain calculations.

A situation in which all possible outcomes are known but not the probabilities is sometimes misleadingly referred to as *strong uncertainty*. Some (e.g., Woodward and Bishop, 1997) perhaps even more incongruously call it 'pure' uncertainty. Even here we presume that we know all relevant outcomes. Is this more realistic than assuming weak uncertainty? Not in any meaningful way. Since we do not really even remotely know the range of possible climate change outcomes, we are not merely uncertain but largely *ignorant* about the future. It is a distinction often not fully appreciated in quantitative fields like economics.

[8] For example, in addition to the effects of policy on carbon emissions, we are uncertain about the effect of the interplay of these on the accumulation over time of these gases; also about the effect of accumulation on mean global temperatures and the effect of this on regional climates; subsequent policy responses; effects on regional wellbeing; etc.

[9] Davidson (1991) prefers the term 'risk' in order to more easily distinguish it from qualitative uncertainty. Yet while useful for modeling and thought experiments, the notion of weak uncertainty is an example of the 'ludic fallacy,' which refers to the confusion of games or math models with real life. Only probabilities in games can be known a priori. The taxonomy presented in Table 1 is slightly different from Keynes's or Knight's notions of uncertainty, and more like that employed by Dovers and Handmer (1995), who emphasize the distinction between ignorance and uncertainty.

Table 1: Classifying uncertainty and ignorance

	Outcomes	Probability Distributions	Possible Outcomes	Principal need
Certainty	Known	Known	Known	na
Weak uncertainty	Unknown	Known	Known	Outcome
Strong uncertainty	Unknown	Unknown	Known	Probability distribution
Informational ignorance	Unknown	Unknown	Unknown	Greater information on critical inputs
Cognitive ignorance	Unknown	Unknown	Unknown	Improved assimilative capacity
Systemic ignorance	Unknown	Unknown	Unknown	Understanding of complex processes

The three classes of ignorance listed in Table 1 are not mutually exclusive, and indeed all three most likely apply here.[10] Informational ignorance is likely the least of the three problems, and the most surmountable. Arguably, we are already at a point where scientific data that are necessary as inputs into climate models – ocean acidity levels, temperatures, icecap melting rates, etc. – are increasingly available or accessible, thanks to ongoing general improvements in measurement techniques. While there remains much that is still not known, progress in this area is beyond doubt.

More problematic are the other two types of ignorance. Cognitive ignorance refers to the natural limit – cerebral or computational – in our abilities to represent reality through modeling, analysis, and calculation, and to parse seemingly unlimited amounts of data and information. In other words, while human ingenuity and information processing capacity are seemingly without limit, there are some problems so complex as to be intractable even for smart humans and powerful computers. Science is known to present such problems; it is a challenge that neuroscientists, despite (or maybe because of) their recent discoveries, are sure to appreciate. So even *if* we obtained accurate measurements of all the relevant variables, it might not be reasonable to assume, given cognitive ignorance, that even our best models would be able to reveal the 'truth.'

We might therefore say that cognitive ignorance involves 'unknown unknowns,' as opposed to the 'known' unknowns dealt with by informational ignorance. Systemic uncertainty, in contrast, covers what I call *unknowable* unknowns, owing to the extreme complexity involved. This refers to conundrums that are potentially beyond the reach of modern scientific inquiry. While cognitive ignorance could in principle be transcended, systemic ignorance might never be fully overcome. While such an idea is likely to arouse controversy,[11] the paradox of moving further from understanding as we acquire more information is not unheard of in science. Modern-day physicists trying to unify quantum mechanics with Einstein's theory of general relativity no doubt appreciate that the progress they make is anything but linear.

[10] Despite its importance, a fourth type of ignorance – 'willful' ignorance – will not concern us here. This relates to the manipulation of information by some for institutional or political economy purposes in a manner that many others remain ignorant. It is similar to what Davies and McGoey (2012) call 'productive' ignorance.

[11] I recognize that this idea is anathema to most scientists, but it is not necessary to argue the point here. More relevant is whether certain unknowns could become 'knowable' near enough in the future to make a meaningful difference in addressing climate change. Here matters appear more dubious.

4. Valuational uncertainty: Subjectivity, circularity, and threshold effects

Compounding the limitations of quantitative benefit-cost assessments is what I call *valuational* uncertainty. It involves B_t from Equation 1, or the value of the benefits that would result from averting climate change damage. I assume, for the sake of simplicity, that *we know* ϕ_t and τ_t – i.e., that we have perfect foresight about what would befall us as a consequence of any possible combination of mitigation and adaptation strategies. Here we are strictly interested in the potential monetary benefits of such policy.

Even if climate scenario ϕ_t and the corresponding adaptation regime (τ_t) were objectively knowable in principle, the value of B_t would not only be unobservable, it could never be objectively ascertained. Quantitative expression of B_t would require monetary assessment of broad changes in human wellbeing as well as valuation of human lives. Such assessments are inescapably subjective.[12]

Some values, to be sure, involve greater uncertainty than others. However imperfect they might be, markets exist for natural resources like minerals or timber, making the monetary valuation of such resources relatively simple. Furthermore, economists have devised methods such as hedonic pricing or contingent valuation for values relating to the environment for which markets do not exist (e.g., endangered species, aesthetic values).

But valuation becomes intractable in the case of ecosystems contributing multiple benefits. How would we, for example, value the nutrient cycling function of the soil or the carbon sequestration function of our oceans? Absent markets, there are no prices, but insufficient understanding of the interrelated nature of the different ecological functions necessarily makes assessment of their value highly subjective with or without prices. Yet the values cannot be ignored; indeed, ecosystems often possess significant values that at least indirectly impact human livelihoods and lives.

Climate change is a particularly thorny problem because of the vast predictive uncertainty discussed earlier. How can we value the damage resulting, say, from *x* additional parts per million of carbon, when we have a poor understanding of secondary and tertiary effects? Moreover, changes are unlikely to be linear and we presently lack sufficient understanding of the relevant ecological thresholds, beyond which predicting outcomes becomes virtually impossible. In this way, predictive uncertainty becomes intertwined with valuational uncertainty – or the question of what values to assign.

But subjectivity also plays an important role in valuational uncertainty. Since future climate change portends some measure of harm to humans in terms of loss of life and adverse health effects, it seems not only reasonable but also imperative that assessed values reflect these damages. The problem is how this can be ensured. The common orthodox economics argument is that, in our society's customary willingness to undertake risky projects, we

[12] Many believe that it is morally wrong to put dollar values on the natural environment or on the impacts of changes to it on human lives. It is more of an ethical stance than a practical argument, one that opposes relegation of humans or nature to the level of marketable commodities. Detractors insist on assigning monetary values to environmental externalities, ecological benefits, and the like, under the rationale that failure to do so implicitly assigns them values of zero, effectively inviting their eventual degradation or overexploitation. The debate goes beyond our present scope since the principal issue here – the valuational uncertainty involved in monetary assessments of the environment and impacts of changes to it on human lives – is not obviated by its resolution.

'reveal' that we attach some finite value to human life or any harm inflicted upon it (else we would never subject ourselves to any nonzero risk, however infinitesimal). Valuation should then merely be a matter of estimating our 'revealed preference' for human life, through hedonic regression or other conventional techniques, and applying the average estimate to the net benefit function in Equation 1.

Yet before doing so, we should note the circular reasoning in the revealed preference claim. We infer that human life possesses finite value from the fact that we are willing to subject ourselves to some statistical risk in our decisions to build, develop, or otherwise alter the natural environment. As is almost invariably the case with economic analytical models, however, the underlying assumption is that decision makers possess perfect information. In other words, agents are making decisions based on a *known* 'life value.' Effectively, we are assuming that we indirectly know that which we are trying to estimate.

In summary, the challenges to accurately assessing the net monetary benefit of climate change policy are immense. Yet even if we *did* have a reasonable sense of what to expect in the future and a practical means of establishing monetary values, there would remain the question of how to account for the value over time. That is, we would need to address the inescapable tradeoff between the net benefits to present versus future generations. This uncertainty is a question of morality or ethics. To it we now turn.

5. Moral uncertainty: The intractable ethics of discounting

The social discount rate and the Ramsey equation

Let us, for the sake of argument, continue assuming that we have perfect foresight regarding ϕ_t and that, in addition, we possess a means of objectively ascertaining the quantitative value of all the damages averted. In this hypothetical world, it would appear that benefit-cost analysis is all that we would need to determine the optimal mix of climate change policies. Yet if we did not consider the social discount rate – '*d* ' in Equation 1 – we would be disregarding the important question of the value of damages in the present compared to those in the future. The issue is central to the controversy over the Stern Review, mentioned earlier. Researchers were quick to note that its primary policy conclusion – the urgency with which climate change should be addressed – hinges crucially on the chosen social discount rate. I label *moral uncertainty* the ethical challenge of what discount rate to employ for climate change policy.

Discounting is a familiar concept when applied to problems of private savings or investment. The discount rate generally represents the pure rate of time preference (a measure of impatience) in the case of the decision of whether to consume or save; in the case of investment, it typically represents the opportunity cost of capital – i.e., the prevailing rate of return on alternative projects.

But applying discounting to projects producing not only private returns but also positive or negative (environmental or social) externalities immensely complicates matters. 'Payouts' – i.e., benefits and costs – are generally experienced by different parties and at different times. Are these parties equally well off to begin with? (Or will they be?) To what extent does this matter? And to what degree should individuals risk uncertain challenges in the future in return

for a greater payoff today? At the core of moral uncertainty is our doubt regarding how much relative importance to accord to future generations.

Some (e.g., Nordhaus, 2007) do not see the problem as too different from a private savings or investment problem, believing that the social discount rate should approximate the private rate of discount. Using the private discount rate for climate change policy *might* make sense if future damage costs were limited to property and other tangible private values as opposed to broader benefits and costs related to human life and wellbeing. But we know that this is not so. Consequently, others (Beckerman and Hepburn, 2007; Broome, 2008) are more circumspect – ambivalent, even – conceding the fundamental ethical nature of the decision. Yet all of the above, and indeed virtually all economists, advocate a positive discount rate in principle.

The dominant belief is that a positive discount rate compensates us for the widespread expectation that future generations will on average be richer than us. But it is not axiomatic that they will be. A high discount rate, in theory, increases the rate of exploitation of the environment and natural resources. Martinez-Alier has noted (1987) that expecting future generations to be richer than us under such circumstances requires assuming an extraordinary elasticity of substitution between the natural environment and whatever might replace it. Also, while we can be optimistic about the advance of scientific knowledge, our optimism does not imply a belief in the discovery of substitutes sufficient to maintain present patterns of consumption in perpetuity, never mind at such a rate to allow for their *growth*, which is what we would infer from a positive social discount rate.

The economic theory behind the social discount rate (d) is generally based on the optimal consumption discount rate suggested in a seminal article by Ramsey (1928), which is based on the tradeoff an economic agent faces between present and future consumption. In the sense in which it shows how to optimize consumption (often taken to be synonymous with wellbeing) over time, his landmark model is suitable for application to the climate change problem. One could argue that it allows us to account for the above concerns, in addition to our expectation regarding the average utility (or income) growth rate into the future. The standard formula for the social discount rate is as follows:

$$d = \delta + \eta g \,, \qquad (2)$$

where d represents the social discount rate, δ stands for the utility discount rate – somewhat akin to the pure rate of time preference or index of 'impatience' – η the elasticity of marginal utility with respect to consumption, and g the expected future growth rate of consumption.

Based on Equation 2, part of the justification for a positive discount rate is a δ that is presumably greater than zero. In other words, on average we value the consumption or wellbeing of today's people more highly than that of people in the future. This might be plausible when considering the consumption-savings decision in general, but is dubious when applied to climate change. Here, the possibility of increased 'negative wellbeing' over time (in the form of an expected wellbeing dominated by low probability but extremely high cost catastrophe events) needs to be considered.

Could we, then, justify a positive δ when evaluating climate change net benefits? It is a philosophical matter not taken up here, but suffice it to say that a positive δ is not universally

accepted. Broome (2008), for one, argues that we should be 'temporally impartial,' and I am inclined to agree. Regardless, there is no way that the choice between discounting future lives and *not* doing so can be anything but subjective.

Things become more muddled when we get to the ηg product in Equation 2. The problem is that while η is meant to represent the elasticity of marginal utility with respect to consumption, such a characterization is so general as to be misleading. As many (Atkinson *et al*., 2009, and Beckerman and Hepburn, 2007) have noted, the elasticity of marginal utility with respect to consumption should itself be decomposed into three distinct indicators, all highly relevant to the discount rate and climate change in general. These are individual risk aversion (the original consumption-savings problem), society's inequality aversion, and society's aversion to *intergenerational* inequality. So among other problems, we have the issue of one of the formula's parameters serving three functions.[13]

As for the expected average growth rate (g), including it as a component of the social discount rate is potentially self-countervailing. True, the assumption that future economic growth will be positive and that future people will therefore be better off than us should call for a positive discount rate. But the potential consequence of a positive rate, especially a very high one – that we stop looking after the needs of future people – undermines the original assumption. If, in other words, we are short sighted today, there will likely be less productive capital available for future use (leading to slower growth).

The ethics of climate change discounting

Ethics aside, then, the social discount rate is fraught with methodological weaknesses. Does this mean we should dispense with it as a policy tool? Beckerman and Hepburn (2007) are among those supporting the use of Ramsey's formulation, despite its faults. They argue that it provides a means of avoiding the two extremes of marketplace 'revealed ethics' – where the social discount rate is inferred from market behavior in relation to private investments – and the dictates of 'philosopher kings' who purport to have some particular expertise and because of this should be entrusted with a potentially momentous decision that could be made arbitrarily.

Broome (2008), in contrast, maintains that reliance on professional expertise is not inconsistent with democracy, as long as the discount rate established by the experts is merely one of many inputs into the (democratic) policymaking process, as opposed to itself being an output computed from shaky inputs based on questionable assumptions and ethical stances. He nevertheless appears to have no issue with expressing our social preference for future generations as a precise figure.

His is a questionable stance, given the inescapable methodological and ethical issues involved. In an article describing secular attitudinal changes in the United States, Oberhoffer (1989) claims that our 'cultural discount rate' has increased over time. Yet he uses the term metaphorically to support his claim that Americans today place less importance on posterity

[13] These authors, among others, have noted the potentially contradictory nature of allowing η to simultaneously reflect these three policy-relevant concepts. A relatively high η, for example, implies greater risk aversion, calling for more aggressive climate change policy. Yet at the same time it reduces the weight of future consumption relative to the present, implying *less* aggressive policy. Atkinson *et al*. (2009) find weak grounds for combining the three concepts into one indicator, and conclude that they should instead be disentangled and represented as three separate variables.

than we once did, nowhere seeking to express it numerically. Oberhoffer appears to appreciate the dangerous seductiveness of a precise figure as a substitute for critical judgment and, to his credit, rejects it.

Broome should adopt a similar tack in the case of climate change. Discounting makes eminent sense for financial investments, but in the social arena it is a misguided attempt to quantify and thereby objectify what is a fundamentally ethical and therefore qualitative problem.[14] Its use is inappropriate for a problem as urgent and momentous as climate change.

There is, in short, no objective basis for ascertaining a 'correct' discount rate on which to base future climate change. Many social benefit-cost studies elide the question of the appropriate discount rate, simply assuming one and burying this detail in otherwise ostensibly objective results. For a high stakes problem like climate change, the enormous moral uncertainty that we have seen is further unambiguous grounds for rejecting social benefit-cost analysis as a policy tool. All three types of uncertainty – predictive, valuational, and moral – argue for extreme caution in drawing conclusions from figures that are based on imperfect understanding of the multidimensional challenge that is climate change.

6. Conclusion

Climate change is a truly interdisciplinary problem, not only involving atmospheric sciences, but also economics, ecology, philosophy, and ethics. The different disciplines introduce different layers of climate change uncertainty; here I have focused on three. *Predictive* uncertainty refers to the difficulty, owing to the extreme complexity of the global climate and attendant feedback effects, in achieving any degree of precision in anticipating future outcomes. *Valuational* uncertainty refers to the impossibility – both because of links to the previous and because of inherent subjectivity – of assigning monetary values to the future damages or to the benefits of preventing them. Finally, *moral* uncertainty refers to the irresolvable matter of how much relative importance to place on the wellbeing of future generations of people.

Given these three forms of uncertainty, it would seem reckless to insist on quantitative precision as a condition for undertaking ameliorative policy. Make no mistake: the extreme caution that we should observe in drawing conclusions from precise figures on highly uncertain matters does not translate to extreme caution in climate change policy. Not knowing the precise costs and benefits is hardly reason to dither, especially since it is well established that climate change is real. Woodward and Bishop (1997) correctly argue that a policy stance that assumes the worst is the most robust to alternative future scenarios.

It brings to mind the precautionary principle, whereby the burden of proof is on agents who would irreversibly – and potentially adversely – alter the natural environment through economic activity. Principle 15 of the 1992 Rio Declaration (UNCED, 1992) states:

> In order to protect the environment, the precautionary principle shall be widely applied by the States [UN members] according to their capabilities.

[14] For far more elaboration on this point, see Best (2012), Schumacher (1973), and Torras and Surie (2016).

> Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation.

While the precautionary principle is usually invoked in reference to a specific project, such thinking is overdue for more encompassing global problems such as climate change.

Social benefit-cost analysis is unreliable as a guide to how aggressively we should address climate change; but it is almost as unreliable in helping us decide upon an appropriate mix of mitigation and adaptation policies. One of the more under remarked realities we confront is that it *might be too late* for mitigation to offer significant hope and that it might make more sense to direct our resources at adaptation and survival. Whether or not this is true depends, of course, on the extent of existential risk that our species faces. We are presently deeply uncertain about this, and it is a subject sure to inspire much future research.

The foregoing discussion hints at two other potential research streams. The first is the possible elaboration of a new paradigm for economic analysis that extends beyond the narrow notion of predictive uncertainty (including its exceedingly narrow treatment even of this) and dispenses with the mathematically convenient assumption of perfect information and foresight. As a policy science, economics must recognize that it is both impossible and undesirable to purge uncertainty, and the subjectivity that often relates to it, from most evaluative contexts. Any future challenge to the dominance of neoclassical economics would do well to consider this.

Second, and following from this, is a greater exploration of policy approaches that more evenly balance quantitative and qualitative concerns. Economists traditionally rely on forecasting, which seeks to reduce all relevant phenomena to numbers in order to more precisely 'predict' the future. If nothing else, the foregoing discussion suggests that a matter as complex as climate change calls for a more 'humble' policy that concedes lack of precision and instead considers scenarios or ranges of possible outcomes. When dealing with a situation with a nonzero probability of catastrophic effects on humans, it is preferable to satisfice based on expert judgment on what we know and what we do not than to seek to optimize based on information that is precise but most likely wrong.

References

Ackerman, F., and C. Munitz (2016). A critique of climate damage modeling: Carbon fertilization, adaptation, and the limits of FUND. *Energy Research and Social Science*, 12: 62-67.

Allen, M., and D. Frame (2007). Call off the quest. *Science*, 318: 582-583.

Atkinson, G., S. Dietz, J. Helgeson, C. Hepburn, and H. Sælen (2009). Siblings, not triplets: Social preferences for risk, inequality and time in discounting climate change. *Economics: The Open-Access, Open-Assessment E-Journal*, 3(26): 1-28.

Beckerman, W., and C. Hepburn (2007). Ethics of the discount rate in the Stern Review on the Economics of Climate Change. *World Economics* 8(1): 187-210.

Best, J. (2012). Bureaucratic ambiguity. *Economy and Society* 41(1): 84-106.

Broome, J. (2008). The ethics of climate change. *Scientific American*, June 2008, 97-102.

Cline, W. (2004). Meeting the challenge of global warming. Copenhagen Consensus Challenge Paper, Copenhagen.

Davies, W., and L. McGoey (2012). Rationalities of ignorance: on financial crisis and the ambivalence of neo-liberal epistemology. *Economy and Society* 41(1): 64-83.

Davidson, P. (1991). Is probability theory relevant for uncertainty? A Post-Keynesian perspective. *Journal of Economic Perspectives*, 5(1): 129-143.

Dovers, S., and J. Handmer (1995). Ignorance, the precautionary principle, and sustainability. *Ambio*, 24(2): 92-97.

Ertürk, K., and J. Whittle (2015). Climate Change, Procrastination and Asymmetric Power. *World Economic Review*, 5: 40-53.

Gowdy, J. (1994). *Coevolutionary Economics: The Economy, Society, and the Environment*. Kluwer Academic Publishers, Boston.

Hansen, J., and M. Sato (2016). Regional climate change and national responsibilities. *Environmental Research Letters*, 11(3): 1-9.

Hodgson, G. (2011). The eclipse of the uncertainty concept in mainstream economics. *Journal of Economic Issues*, 45(1): 159-175.

Hope, C. (2008). Optimal carbon emissions and the social cost of carbon under uncertainty. *The Integrated Assessment Journal*, 8(1):107-122.

Martinez-Alier, J. (1987). *Ecological Economics*. Blackwell, Oxford and Cambridge.

Morgan, J. (2016). Paris COP21: Power that speaks the truth? *Globalizations*, http://dx.doi.org/10.1080/14747731.2016.1163863

Nordhaus, William (2007). A review of the Stern Review on the economics of climate change. *Journal of Economic Literature*, 45: 686–702.

Nordhaus, W., and J. Boyer (2000). *Warming the World: Economic Models of Global Warming*. Cambridge: MIT Press.

Oberhofer, T. (1989). The changing cultural discount rate. *Review of Social Economy*, 47(1): 43-54.

Pachauri, R., A. Reisinger, Eds. (2007). *Climate Change 2007 Synthesis Report*. Intergovernmental Panel on Climate Change.

Ramsey, F. (1928). A mathematical theory of saving. *Economic Journal* 38: 543-559.

Rayner, S. (2012). Uncomfortable knowledge: the social construction of ignorance in science and environmental policy discourses. *Economy and Society* 41(1): 107-125.

Rosen, R., and E. Guenther (2014). The economics of mitigating climate change: What can we know? *Technological Forecasting and Social Change*, 91: 93-106.

Schumacher, E.F. (1973). *Small is beautiful: Economics as if people mattered*. New York: Harper and Row.

Spash, C. (2016). The political economy of the Paris Agreement on human induced climate change: A brief guide. *Real-World Economics Review* 75: 67-75.

Stern, N. (2007). *The Economics of Climate Change: The Stern Review*. Cambridge University Press, 712 pp.

Stern, N. (2013). The Structure of Economic Modeling of the Potential Impacts of Climate Change: Grafting Gross Underestimation of Risk onto Already Narrow Science Models. *Journal of Economic Literature* 51(3): 838-859.

Tol, R. (2003). Is the uncertainty about climate change too large for expected cost-benefit analysis? *Climatic Change* 56: 265-289.

Tol, R. (2009). The economic effects of climate change. *Journal of Economic Perspectives*, 23(2): 29-51.

Torras, M., and G. Surie (2015). On the quantitative bias in economics. *Journal of Economic Issues*, 49(4): 1028-1044.

UNCED (1992). *Rio Declaration on Environment and Development*. United Nations Conference on Environment and Development, United Nations Environment Programme.

Weitzman, M. (2009). On modeling and interpreting the economics of catastrophic climate change. *The Review of Economics and Statistics*, 91(1): 1-19.

Woodward, R., and R. Bishop (1997). How to decide when experts disagree: Uncertainty-based choice rules in environmental policy. *Land Economics*, 73(4): 492-507.

Author contact: torras@adelphi.edu

SUGGESTED CITATION:
Mariano Torras, "Uncertainty about uncertainty: the futility of benefit-cost analysis for climate change policy", *real-world economics review*, issue no. 77, 10 December 2016, pp. 11-25,
http://www.paecon.net/PAEReview/issue77/Torras77.pdf

Growthism: a cold war leftover

Herman Daly [University of Maryland, USA]

Those of us old enough to remember the Cold War will also remember that it involved a growth race between Capitalism and Communism. Whichever system could grow faster would presumably win the allegiance of the uncommitted world. The idea of a steady-state economy was therefore anathema to both sides. The communist growth god failed first because of stifling political repression and economic inefficiency. But the capitalist growth god is now failing as growth becomes uneconomic due to compounding environmental and social costs, and is propped up only by fraudulent accounting, monopoly, and financial corruption. Neither system can accept the idea of a steady-state economy, but neither can attain the impossible alternative of continuous growth.

Ecological economists and advocates of a steady-state economy are accustomed to attacks from capitalists, which have by no means disappeared. We are less accustomed to attacks from the left, not from communists who have virtually disappeared, but from remaining Marxists and socialists. Although Marxism is largely discredited (along with other manifestations of 19th-century determinism, such as Freudianism and Eugenic Darwinism), one cannot by any means take that as a vindication of capitalism, which has only gotten worse in its quest for unending growth. In spite of my distrust of Marxism, there are some "green Marxists" who are allies in that they, unlike either neoclassicals or traditional Marxist economists, have seen that growthism is the big problem for us all (e.g. John Bellamy Foster, Brett Clark, Richard York, *The Ecological Rift*).

Recently, another socialist (I am not sure to what extent he considers himself a Marxist) has criticized the steady-state economy for being essentially capitalist. This is economic historian Richard Smith. He sees the steady-state economy as a distraction from the true solution to growthism, which he claims is "eco-socialism".

One should be grateful to one's critics–it is much better to be criticized than ignored. Richard Smith kindly takes me as exhibit A for a position that he misleadingly labels "steady-state capitalism". I have never used that term, always speaking of a steady-state *economy*, which is neither capitalism nor socialism, although it draws features from both. Indeed, in the Cold War context it was thought to offer a Third Way, a possibility for uniting the best features of each system. Change is impossible unless you start from where you are. Just abandoning "capitalism" and proclaiming, "let there be eco-socialism" is not a solution to the error of growthism. However, it is at least a refreshing change to be attacked, and on balance rather politely, by a socialist who, unlike many neoclassical-Keynesian growthists, has taken the trouble to learn about the steady-state economy. Disagreements will follow, but my appreciation for Smith's critical attention needs to be expressed.

Richard Smith characterizes capitalism as a system that must "grow or die". It then follows immediately that since capitalism must grow, it cannot be a steady state. OK then, if capitalism cannot be a steady state, then neither can a steady state be capitalism. So let's not speak of "steady-state capitalism". I, for one, never have – although Mr Smith tendentiously

attributes that term to me. By the same logic, following Marx, one might define socialism as a classless society based on overwhelming material abundance arrived at through rapid economic growth under the centrally planned dictatorship of the proletariat. Socialism also depends on growth. Therefore steady-state socialism is impossible. It was precisely to avoid such sterile definitional disputes that I always said "steady-state *economy*", and never "steady-state *capitalism*", or socialism for that matter. Recently I see that Smith has used the term "green capitalism" to identify his target. I think that is much more accurate.

I think it would be more productive to start by defining a steady-state economy, followed by ecological and ethical arguments for its necessity and desirability. We could then avoid ideological classifications based on abstract definitions of what capitalism or socialism "essentially must always be". We now live in a full world. Capitalism and socialism are both from the empty-world era in which growth was the *desideratum*. Must we insist on pouring new wine into old wineskins, and then watching them burst?

Smith's unhappiness with me derives most specifically from my preference for the market over centralized planning as a tool for dealing with the single technical problem of allocative efficiency. Steady-state economics deals with three problems: sustainable scale, just distribution, and efficient allocation. It takes the first two issues, scale and distribution, *away from the market*. It calls for quantitative ecological limits on the throughput of resources so that the market can no longer determine the physical scale of the economy relative to the biosphere. It also advocates social limits to the range of income inequality, so that the market can no longer generate large inequalities of wealth. Subject to these two major and prior macro-level constraints, it then relies on the market to efficiently allocate resources. This is not advocacy of the Market with a capital M, the deified master evaluator and controller of life. This is market with a small m, a limited tool for rationing resources, communicating information, and exchanging goods and services. To understand the market in this limited sense as a means of solving the allocative problem, one, especially a socialist, still cannot do better than to read the 1936 classic, *On the Economic Theory of Socialism*, by Polish economist Oskar Lange (with Fred M. Taylor).

Reliance on markets for allocation (now within prior ecological and distributional limits) is further constrained, even within traditional microeconomics, by opposition to monopoly, and restriction of market allocation to rival and excludable goods. Non-rival and non-excludable goods have long been recognized to require non-market allocation. Even so, Mr Smith is still unhappy with any role for markets.

Richard Smith deserves credit for recognizing and opposing the real evils of financial-monopoly-crony capitalism as it currently exists. And, unlike both traditional Marxists and neoclassical economists, he realizes that we cannot grow forever, and that we have in many dimensions already far overshot optimal scale. And he takes the trouble to debate critical issues rather than ignore them. However, he thinks only socialism can somehow cure these evils. The operative word here is "somehow". Somehow we must wipe the slate clean of any institutions associated with markets, such as property, division of labor, exchange, and profit. How? By violent revolution? By rational persuasion? By moral conversion? That is left vague. It is all very well, for example, to point out the real problems with excess reliance on the profit motive. But if we abolish profit as a source of income then we also abolish self-employment. Everyone must then become an employee earning a wage. Who then is the employer? Do we all then work for Ajax United Amalgamated Corporations? Or for the Universal State Monopoly? Is there something about the mere act of market exchange, and the category of

profit, (not just excessive inequality and monopoly ownership of the means of production) that offends or confuses Marxists?

Nevertheless, if Marxists now advocate limiting growth, that is a *big* and welcome change. Maximizing growth to achieve overwhelming material abundance has been seen as the path to the "new socialist man", who, according to Marx, can only be freed from his bourgeois greed by objective abundance, by the abolition of scarcity, not by the "utopian" morality of sharing. I have never seen a Marxist proposal to limit the scale of the macro economy to an ecologically sustainable level – nor for a maximum as well as a minimum income to limit the range of distributive inequality to a reasonable and fair degree. Rhetorical calls for absolute equality and abolition of private property abound, but are neither realistic nor fair.

Marxists also go far out of their way not to recognize overpopulation and the need to limit population growth – a critical dimension of both scale and distributive inequality, given class differentials in fertility and in access to contraception. (I am glad to note that Richard Smith is apparently an exception to this leftist tendency). A stationary population is part of the definition of a steady-state economy.

A limited range of income inequality would restrict the ability of the rich to bid necessities away from the poor in the market. Unjust distribution of income does get reflected in markets, but let us attack the unjust reality, rather than break the mirror that reflects it. Furthermore, quotas on basic resource throughput could raise prices enough to eliminate most frivolous and wasteful production, as well as stimulate recycling, and increase efficiency while ruling out the Jevons effect. If we start with depletion quotas on basic resources, then the resulting increase in resource prices and efficiency cannot lead to more use of the resource. Auctioning transferrable quotas rather than giving them away (*markets* rather than direct government allocation, *pace* Mr Smith) will raise enough revenue to greatly reduce taxes on the poor. The auction would capture the increased resource rents for redistribution in good Henry George fashion.

It is not at all clear why Smith thinks markets must always be bad masters rather than good servants. If we forgo markets, should we then perhaps have another go at central planning and collectivization of agriculture? Would Mr Smith have preferred War Communism to Lenin's New Economic Policy because the latter was really just "state capitalism" that re-established significant reliance on markets? To be fair, we do not know what Smith thinks about any historical experience with the abolition of markets because he does not mention any.

If "eco-socialists" reject the steady-state economy as "inherently capitalistic", then what specific policies do they recommend? How do their policies differ from those of steady-state economics? Are there some policies we agree on?

Both steady state and eco-socialist critics of growthism should be united in humility before a common dilemma – namely that the bought-and-paid-for government that would have to enact the programs needed for a steady-state economy is the same government that would have to run a socialist economy. A government that cannot conceive of nongrowth, that cannot break up too big to fail monopolies, or provide debt-free money as a public utility, or tax carbon, will certainly not be able to administer a centrally planned economy – nor a steady state, even with the help of markets. We have deeper problems of moral and spiritual renewal (in addition to recognition of finitude and the laws of thermodynamics) that transcend both

capitalism and socialism. Some of these problems were identified better in Pope Francis' *Laudato Si* than in any economics journals that I am aware of. In today's secularist world it is admittedly hard to envision the source for the basic moral renewal required to face growthism's threat to both sustaining and sharing the Earth's capacity to support life. Marxist dialectical materialism and collectivism seem to me already to have historically demonstrated their failure in this regard. Nor does the intelligentsia's current religion of Enlightenment Post-Modernism offer much guidance. We need something more basic and fundamental. Although things look bleak, we never know enough to justify giving up hope. But we should try to avoid repeating past mistakes.

Author contact: hdaly@umd.edu

SUGGESTED CITATION:
Herman Daly, "Growthism: a cold war leftover", *real-world economics review*, issue no. 77, 10 December 2016, pp. 26-29, http://www.paecon.net/PAEReview/issue77/Daly77.pdf

Finance capital and the nature of capitalism in India today

C. P. Chandrasekhar [Jawaharlal Nehru University, New Delhi, India]

A distinctive feature of the current phase of capitalism is the growing role of finance capital as an instrument to establish capitalist hegemony and facilitate the appropriation of surplus. In the developed countries this transition towards financial hegemony was the result of the inflationary crisis and stagnation that affected the OECD countries after the late 1960s. The contractionary fiscal and monetary policy response to inflation intensified the deceleration in real growth. This together with the low real interest rates on bank deposits adversely affected banking business and profits. These developments triggered a process of deregulation and liberalisation in the financial sector, ostensibly to allow banks the freedom to reduce losses and find new means to profit.

The deregulation led to an expansion and diversification of the financial sector. With the burgeoning of finance, involving the emergence of new markets and transformation of old ones, creation of new non-bank financial institutions, and the "manufacture" of new financial products, the financial sector was able to increase the volume of its surplus. Securitisation, by making credit assets tradable, allowing for the transfer of risk, and freeing the creators of credit of the underlying risk encouraged a massive expansion in the volume of credit provided. This required expansion of the universe of borrowers, who being driven to borrow, often became willing victims of the drive to transfer surpluses from their incomes directly to financial firms.

Available figures do point to galloping growth in the global operations of financial firms. One obvious form this has taken since the international lending boom of the late 1970s is the expansion of operations of international banks in less developed countries, especially the so-called "emerging markets". At the time of the East Asian crisis (mid-1997), the international asset position of banks resident in 23 countries reporting to the Bank of International Settlements stood at $9.95 trillion, involving $8.6 trillion in external assets after adjusting for local assets in international currencies. By June 2007, before the global financial crisis that precipitated the Great Recession, when 40 countries were reporting, this had risen to $33.71 trillion, with external assets totalling $29.98 trillion. This trend characterised countries that reported on both dates as well. For example, the international assets of UK-based banks had increased from $1.5 trillion to $6.1 trillion, and that of US banks from $0.74 trillion to $2.8 trillion.

One consequence of the post-1970s expansion of liquidity in the international financial system was the need on the part of international finance capital to find new avenues to lend and invest. Having to keep money moving to earn returns, and running out of options within the developed world, private international finance that had excluded most developing countries from its ambit because they were perceived to be too risky both economically and politically, chose to target some developing countries that were soon identified as emerging markets. Suddenly, flows of private financial capital to developing countries, which till then had access

to foreign capital only in the form of limited flows of foreign direct investment and "aid' from the bilateral and multilateral development aid network, became a possibility, with an implicit message that this was available on demand.

To exploit this option, developing countries needed to dilute controls that had been imposed on flows of foreign capital, especially foreign financial capital wanting to enter their equity, debt and insurance markets. With hindsight we know that almost all developing countries chose to exploit this option at different points of time starting in the 1970s in the Southern Cone countries of Latin America. The reasons differed across countries. To fathom the explanation for those differences, we need to turn to the dissimilar developmental strategies that were adopted by developing countries in the aftermath of the period that saw two World Wars, an agricultural depression, the Great Depression, and decolonisation. In some countries the wars and the movement against colonialism led to the institution of governments led by socialist forces that opted for development within the framework of central planning. But most countries, including India, came under governments that opted for a capitalist path of development, starting from a situation where their societies were characterised by substantial semi-feudal remnants.

Among those that opted for a capitalist path, most countries, on the basis of their experience with being predominantly agricultural producers and exporters of primary products, chose to adopt a strategy of import substituting industrialisation. Pursuing import-substituting industrialisation required the strengthening of indigenous industry, not just with protection, but by control and regulation that restricted the role of large and predatory foreign capital and disciplined domestic business to behave in ways that served national development. The degree to which this was done varied across countries. For historical reasons, though the Indian State represented an alliance of the domestic bourgeoisie and landlords, the Indian government adopted a development strategy that involved substantial state intervention. This included, besides controls on cross-border flows of capital and domestic regulation of capacity creation, production and prices, wide public ownership and an emphasis on "planning" for successful economic development, along lines being pursued at that time in the Soviet Union.

Thus, even among underdeveloped countries launching on development after World War II, India was in many senses unique. In terms of choice of the mix of emphasis on industrialisation based on the domestic market and that driven or facilitated by exports it was focused almost wholly on an internally oriented growth strategy. This did seem warranted for a number of reasons. At independence in 1947, India was a country that showed much promise as a potential candidate for successful industrial development. It already had considerable experience with factory production, with the first successful factory established within the country dated to 1854. In the event, after the Second World War, India was among the more industrialised of the underdeveloped countries, especially those that had had just come out of colonial domination. This historical legacy did favour India when the development experiments of the post-War, decolonisation years began.

Further, despite the low level of its per capita income in India at that time, the sheer size of the country as defined by its geographical area and its population meant that it had a reasonably large sized market for manufactures. Moreover, the substantially unequal distribution of income ensured that there existed a significant number of people with income levels that implied a rather diversified demand for manufactures. Even though this section

was proportionately small, the large size of the population made their numbers significant. These factors provided the foundation for a reasonably large and diversified domestic market.

However, growth based on the domestic or home market was not easy to realise and had been achieved in practice by very few countries since the Industrial Revolution. A prerequisite for sustained growth was the growth of the mass market for manufactures. This required land reform that broke down semi-feudal relations in the agricultural sector, so as to accelerate productivity and output growth by providing the actual cultivators – whether tenants or peasants – with the means and incentive to invest and by distributing more equally the benefits of that growth so that those benefits could translate into demand for manufactured goods that have a mass market that can support industrialisation.

Ensuring these prerequisites was an extremely difficult proposition, given the alliance between the capitalists and landlords. Not surprisingly, many countries, often after initial attempts at import substituting industrialisation, soon turned to the export market as a potential alternative stimulus to growth. A few sought to do this on the basis of investment by indigenous capitalists (Republic of Korea and Taiwan) supported by developed country governments, while others attempted it by attracting foreign investment that would use local labour reserves and establish capacities aimed at production for the international market. Among these the only underdeveloped country of reasonable size among the delayed late-industrialisers that managed to achieve developed country status was South Korea. Some, like the newly industrialising countries of Southeast Asia in the 1980s and after, experienced rapid growth for significantly long periods and registered a rise in the share of manufacturing production and in per capita incomes. But those years of 'miracle growth' did not last beyond the 1990s and were unable to deliver a South Korea-type transition.

All countries that adopted export-led strategies, despite their different growth trajectories and degrees of success, remained dependent upon and subordinate to foreign capital. In the case of those that had experienced export success, this was partly because they were under pressure to open their financial borders as a quid pro quo for continued access to markets in the developed capitalist countries on which they were dependent. In others, dependence on foreign capital with control over international markets was needed for export success. As a result, in course of time those choosing export-oriented strategies had to open their international economic borders by diluting or dismantling capital controls to allow for inflows of foreign investors seeking new investment avenues in the Age of Finance.

Thus, development that strengthened the political independence gained by countries after World War II by ensuring independence from predatory foreign capital was only possible in countries that ensure successful growth based on the domestic market. So it was significant that at independence India saw the accession to power of a government that had the social sanction needed to formulate and implement a domestic market-oriented national development strategy. Its emergence out of a national movement against colonial rule gave it that sanction.

However, assessed merely in terms of rates of growth, the success of India's post-independence development strategy is partial at best. The initial dynamism, especially of industry, displayed during the decade and a half after 1951, gave way to a period of secular stagnation that stretched between 1965 and the late 1970s. The most obvious indicators of poor economic performance are the inadequate diversification of India's production structure away from agriculture to manufacturing, and the rather premature and rapid diversification

into services that has occurred in recent decades. The share of manufacturing in GDP did rise from around 9 per cent in 1950-51 to 13 per cent in 1966-67. But it did not cross the 14 per cent mark for a little more than a decade after that, and touched 16.4 per cent at its peak in 1996-97. The contribution of manufacturing to employment was even more dismal.

India's experience was in fact worse than that in other similarly placed developing countries. In 1960, industry contributed 37 per cent of GDP in Brazil, 45 per cent in China, 19 per cent in India, 19 per cent in Indonesia, around 25 per cent in South Korea, 19 per cent in Malaysia and 19 per cent in Thailand. By 1985, the figures were 45 per cent in Brazil, 43 per cent in China, 26 per cent in India, 36 per cent in Indonesia, 39 per cent in South Korea, 39 per cent in Malaysia, and 32 per cent in Thailand. Thus, the 1960 to 1985 period was one in which in most developing countries rapid diversification in favour of manufacturing was occurring, but India had not shown the same tendency. The long-term, slow growth and subsequent near-stagnation of the share of industry in GDP in India was more the exception than the rule among developing countries.

Overall, a number of features of India's post-independence growth strategy structurally limited the potential of the system. To start with, despite talk of land reform, of providing "land-to-the-tiller" and of curbing the concentration of economic power, little was done to attack and redress rural asset and income inequality. And while the worst forms of monopolistic practices were curbed, asset concentration in the industrial sector was never really challenged. Rather, India's big business groups were able to use state intervention as a device to consolidate and expand their monopolistic positions. One consequence of the persistence of asset and income inequality was that there were definite limits to the expansion of the market for mass consumption goods in the country. The large mass of peasantry, faced with insecure conditions of tenure and often obtaining a small share in the outputs they produced, had neither the means not the incentive to invest. The prospect of increasing productivity and incomes in rural India, which was home to the majority of its population, in order to stimulate domestic demand was therefore limited.

Under these circumstances, a continuous growth in state spending was essential for the growth of the market. In the event, the stimulus to growth during the early post-independence years came from the state itself. It provided domestic capitalists with a large once-for-all market for manufactures by widening and intensifying protection and displacing imported goods from the domestic market. It sought to expand that market through its current and capital expenditures and it supported the domestic capitalist class by investing in crucial infrastructural sectors. This strategy did pay dividends during the decade-and-a-half immediately following independence when rates of industrial growth were creditable by international standards and India built up a diversified industrial base, and the public sector expanded rapidly to provide crucial infrastructural services, industrial raw materials and capital goods to sustain industrial growth even when the foreign exchange available to import these commodities was limited.

This growth was stalled because of the second of the contradictions characterising the process of development. The State within the post-independence economic policy regime had to simultaneously fulfil two different roles that were incompatible in the long-run. On the one hand it had to maintain growing expenditures, in particular investment expenditure, in order to keep the domestic market expanding. At the same time, however, the State was being exploited as an instrument for the "primary accumulation of capital". Through tax avoidance and evasion, subsidies and transfers, and through State-contracts, private fortunes were

being built up at the expense of the State exchequer. By the mid-1960s, not only was the once-for-all stimulus offered by import substitution exhausted, but the ability of the State to continue to provide the stimulus to growth was also undermined. In the event growth decelerated leading to the "secular stagnation" of the late-1960s and 1970s. Rather than reversing this by undertaking the absent structural and fiscal initiatives that were responsible for the secular stagnation, Indian big business and the State began to look for an opportunity to exit from this strategy and shift to one in which exports would provide the stimulus to growth.

The easy way to pursue that alternative was to imitate the East Asian, second tier industrialisers and attract foreign investment that used the country concerned as a location for production for the world market. But since such investment was dependent on the decisions made by foreign investors, a strategic adoption of export led development had to be supported also by measures to restructure the capacities of domestic agents and make them internationally competitive. This required substantial liberalisation of trade and foreign investment policies, that would immediately lead to an increase in imports, not least because potential exporters would choose to import the technology, capital goods, intermediates and components needed for export production. In practice, a policy of trade liberalisation was adopted on the grounds that the competition it would unleash would help restructure domestic economic activity, render firms and other economic agents in India internationally competitive, and put the country on an outward-oriented, export-led growth trajectory. Even if this does prove to be the "ultimate" result of such trade liberalisation (which it normally is not), this cannot be its immediate fall-out. Restructuring domestic capacity takes time as does the process of finding customers and building 'goodwill' in global markets. On the other hand, post-liberalisation, the pent-up demand, especially of the rich, for imported goods that had thus far been curbed with protection would be immediately released. This would lead to a widening of the trade and current account deficit in the balance of payments of the liberalising economy, with foreign exchange expenditures rising much faster than foreign exchange earnings. So access to foreign capital to finance that deficit is a prerequisite for "successful" liberalisation that is not aborted by a balance of payments crisis.

Thus, the transition to a liberalised, open economic regime could be stalled by the actual and potential balance of payments difficulties associated with that experiment. It was in this context that the access to foreign capital ensured by the rise to dominance of finance was seen as an opportunity. Ensuring access to foreign capital flows resulting from the accumulated liquidity in the international market required in the first instance relaxation of controls on capital inflows, including inflows of purely financial capital into debt and equity markets. But attracting such inflows also requires attracting the carriers of such capital, viz., large financial firms such as banks, institutional investors, pension funds and insurance companies. This required relaxation of the terms of entry into and operation in domestic markets of such firms. It also required changing the regulatory framework in keeping with international norms and guidelines, such as those formulated by the Basel Committee. A shift from the "structural regulation" of the financial sector and financial institutions, to market mediated regulation was the result.

The transformation brought about by the new financial framework had many features. To start with, banks extended their activity beyond conventional commercial banking into merchant banking and insurance. Second, within banking, there was a gradual shift in focus from generating incomes from net interest margins (or the difference between deposit and lending rates) to obtaining them in the form of fees and commissions charged for various financial

services. Third, related to this was a change in the focus of banking activity as well. While banks did provide credit and create assets that promised a stream of incomes into the future, they did not hold those assets till maturity any more, as they used in the past in the so-called "originate-and-hold" model. Rather they bundled them into pools, attached those bundles to particular securities eligible for the stream of incomes due from the underlying assets, and sold these securities for a fee to institutional investors and portfolio managers. Banks transferred the risk for a fee, and those who bought into the risk looked to the returns they would earn in the long term. This was the "originate and distribute" model of banking. It meant that those who originated the credit assets tended to understate or discount the risks associated with them. Moreover, since many of the securities created on the basis of these credit assets were complex derivatives, the risk associated with them was difficult to assess. The role of assessing risk was given to private rating agencies, which were paid to grade these instruments according to their level of risk and monitor them regularly for changes in risk profile. Fourth, the ability of the banking system to "produce" credit assets or financial products meant that the ultimate limit to credit was the state of liquidity in the system and the willingness of those with access to that liquidity to buy these assets off the banks. Within a structure of this kind periods of easy money and low interest rates increased the pressure to create credit assets and proliferate risk. Finally, financial liberalisation increased the number of layers in an increasingly universalised financial system, with the extent of regulation varying across the layers. Where regulation was light, as in the case of investment banks, hedge funds and private equity firms, financial companies could borrow huge amounts based on a small amount of own capital and undertake leveraged investments to create complex products that were often traded over the counter rather than through exchanges. Credit risk transfer neither meant that the risk disappeared nor that some segments were absolved from exposure to such risk.

These changes made the financial sector an important site for profit appropriation. There are a number of stylised facts that support that argument. The first is the sheer size of the financial sector and the growing importance of finance in the growth of national income. The second is evidence of financial over-development with the ratio of financial assets to GDP and of financial assets to real wealth rising sharply. And a third is the rising share of financial profits in total corporate profits. All these are indicators of an accelerated expansion of financial activity as the principal site for surplus appropriation.

There have been other significant consequences associated with the rise of finance. One is a change in the mode of appropriation of surplus by finance itself. In the past finance acquired a share of the surplus generated in the sphere of production of goods and services. Through investments in or loans provided to these activities, finance received a dividend, a capital gain when it exited from ownership and interest on credit provided, which after taking out of its costs determined its net profit. In fact, a lot of financial 'investment' was in equity that afforded it control or influence over corporations engaged in production and/or service provision. Dividend incomes rather than capital gains were the main source of return to financial interests that wanted to retain control of profitable real assets. Matters have changed in recent decades. Further, an expansion in the volume of financial transactions allows for periods or episodes of asset price inflation, which when 'marked to market' (or valued at prevailing market prices) in their books seem to deliver profits and enhance wealth. Not being the result of the sharing or direct appropriation of surpluses generated in the sectors producing commodities and services, this profit and the wealth increase is, at one level, notional. But so long as financial assets are liquid in the sense that they can be easily

encashed and the value of money is protected, this wealth amounts to purchasing power and a means of command over real resources.

A second consequence is a tendency for economic activity outside the financial sector to be shaped by finance. Which sectors turn out to be the sunrise sectors in the economy, which firms flourish and which survive and grow, which shut down or are merged or amalgamated with others, and which 'technologies' tend to get showcased are shaped, often unbeknownst to us, by finance. One illustration of this is the growing importance of 'start-ups' in India's 'new' capitalism. Certain firms experimenting with certain technologies are chosen as the target for financial bets. Few of these survive, even if they are extremely well funded with venture capital in their early years. The international experience shows that even those that survive and deliver huge returns to shareholders and generate fortunes for their promoters, like Netscape, AOL and Yahoo did, fade away, unlike a few others like Google and Apple, which survived and became leading firms. Many others of varying sizes merge or are the subject of takeover. But the gains made from the few successes are more than enough to wipe out the losses associated with the many failures.

A third consequence is our perception of technology itself has changed. Technology in the era prior to the Age of Finance was largely a combination of 'hardware' and 'software' that used a certain set of specified capital goods, intermediates and components to undertake a planned production routine to yield a product with a specific design, technical characteristics and use value within a defined organisational framework, like the factory. This allowed us to breakdown technology into segments such as materials technology, manufacturing technology, design technology, and managerial technology. The last was clearly far less of a technology than the others. But technological change could involve improvements in any of these. A feature of technology in recent years is the growing importance of "software" elements and managerial technology in the spectrum. Today's 'technology' majors include the likes of Google (a search engine), Facebook (a social media platform) and Uber (an aggregator). This allows for both the widening of the scope of innovation and an increase in the pace of obsolescence of technologies, providing a constant source of 'new, new things' on which finance can place its bets.

Fourth, the period of the rise of finance has seen a substantial expansion of the service economy and the GDP generated by services. This is because of the specific way in which finance has moulded the use of information technology, allowing it to transform industries delivering products such as goods, media and music, and access to traditional services like commercial taxi services. One often noted feature of contemporary Indian capitalism, is that recent growth in national income has become overwhelmingly dominated by the services sector, and that too by activities that in themselves do not promise much for future growth increases. What needs to be noted is that this is not the result of the observed diversification of national economic activity in the direction of services that occurs at relatively high levels of per capita income after a period of diversification away from agriculture to manufacturing.

Finally, the rise of finance changes the dynamics of capitalism itself. Some economies still remain exporters, others are the destination for imported profits from foreign investment, and yet others grow on the basis of internal stimuli, in which tax and debt financed public expenditure is replaced with debt financed private expenditure as the principal 'exogenous' stimulus to growth. One cause for this is that for a host of reasons, ranging from its fear that excess borrowing would spur Inflation that would erode the value of the assets or the command over real resources of those holding financial wealth to its desire to rein in States

pursuing proactive fiscal policies based on borrowing, which legitimises the State and delegitimises the market to the detriment of finance, capital opposes debt financed spending by governments. As the presence and power of finance increases, therefore, fiscal conservatism becomes the norm and austerity a recurrent policy recommendation.

All of this suggests that the growing dependence on foreign finance capital has distorted India's growth. India's failure, visible from the mid-1960s, to make an expanding and technologically dynamic industrial sector an important, let alone the principal, driver of growth has only worsened in the age of finance. As per the new series on national accounts with 2011-12 as base, over the entire four-year period starting 2012-13, services of various kinds accounted for as much as 68.7 per cent of total GDP growth. Manufacturing provided for only around 18 per cent, slightly less than the service activities comprising trade, repair, hotels and restaurants. What is more startling is that some of the biggest contributions to GDP growth came from finance, insurance and real estate (FIRE – 30.9 per cent) and public administration and defence (12.5 per cent). Indeed, these two sectors together accounted for 43 per cent – or nearly half – of all estimated increases in economic activity in the past four financial years. Over 2015-16, this FIRE sector accounted for a huge 21.6 per cent of total GDP (or Gross Value Added at Basic Prices, as the new series describes it). This is problematic, because expansion in these sectors is not suggestive of a good foundation for future stable growth.

Accompanying this is the evidence from the demand side that indicates that the combination of financial liberalisation and the large financial inflows that followed did create a new regime of accumulation in India. The inflow of foreign capital had as its counterpart an increase in the overhang of liquidity in the domestic economy. Based on that overhang, a liberalised banking system has been creating new credit assets at a rapid rate. The ratio of bank credit outstanding to GDP, which had remained at around 22 per cent for a decade starting 1989-90, began to rise after 1999-2000, doubled by 2005-06 and is currently well above 50 per cent. India has been witnessing a credit boom during its high growth years.

There were also significant changes in the sectoral distribution of credit. Overall there were two sets of sectors that gained in share. The first comprised of retail advances, covering housing loans, loans for automobile and consumer durable purchases, educational loans, and the like. The share of personal loans increased from slightly more than 9 per cent of total outstanding commercial bank credit at the end of March 1996 to close to a quarter of the total more recently. The second area of change was the distribution of credit going to industry, which at around 40 per cent of total bank credit outstanding was still substantial. The share of infrastructural lending in the total advances of scheduled commercial banks to the industrial sector rose sharply, from less than 2 per cent at the end of March 1998 to 16.4 per cent at the end of March 2004 and as much as 35 per cent at the end of March 2015. That is, even as the volume (though not share) of lending to industry in the total advances of the banking system has risen, the importance of lending to infrastructure within industry has increased hugely. Sectors like power, roads and ports, and telecommunications have been the most important beneficiaries. For commercial banks, which are known to prefer lending for short term purposes, this turn to lending to infrastructure was a high risk strategy.

These changes initially spurred growth because of the demand increases it financed. But soon it was clear that this trajectory was one that involved driving growth at the expense of financial stability, since many of these projects and loans were not viable and turned non-performing. According to the end-June 2016 edition of the Reserve Bank of India's (RBI's) biannual Financial Stability Report, gross non-performing assets (GNPAs) of the scheduled

commercial banks (SCBs) rose sharply from 5.1 per cent of gross advances at the end of March 2015 and to 7.6 per cent at the end of March 2016. Thus, the problem is not just the volume of bad assets, but the rapid growth in such assets. According to answers to two questions (Nos. 1759 and 2526 of August 2016) in the upper house of the Indian parliament, while the total GNPAs of public sector banks stood at Rs. 4,768 billion at the end of March 2016, the non-performing assets that were reported by them in the second half of financial year 2015-16 alone amounted to Rs. 2,770 billion. Figures obtained by Reuters through a Right to Information application indicate that stressed assets on the books of the banks had risen from Rs. 8060 billion at the end of December 2015 to Rs. 9220 billion at the end of June 2016. That would suggest that the value of loans that could turn bad is still on the rise.

Even ignoring this trend, the rise in non-performing assets due to reclassification is not without implications for the health of India's predominantly public banking system. Once assets are recorded as non-performing, banks need to write off loss assets. They must also provide for the implicit decline in the value of doubtful and sub-standard assets. That adversely affects the profitability of banks. Even though much less than the RBI mandated 70 per cent of NPAs have on average been provided for by Indian SCBs, the return on assets (RoA) and the return on equity (RoE) of the group fell between end-March 2015 and end-March 2016, from 0.8 to 0.4 per cent in the case of the former and from 9.3 to 4.8 per cent in the case of the latter. Underlying this profit squeeze was an 86 per cent year-on-year growth of risk provisions and a 27.3 per cent increase in write-offs, which together contributed to a 43 per cent fall in profits after tax. Given the uneven distribution of this hit across banks, 21 SCBs accounting for 37 per cent of the total assets of all SCBs recorded negative RoA values over financial year 2015-16.

The question that arises, therefore, is the manner in which the government, the RBI and the banks are aiming to address this problem. One option would be for the banks to treat the write-offs as merely "technical' and then try and recover as much of the value of these assets as possible, to strengthen their financial position. However, the experience here has been disappointing. Not only has total NPA reduction been flat between 2014-15 (Rs. 1,270 billion) and 2015-16 (Rs. 1,280 billion) when the sum of declared NPAs was rising, but much of this reduction has been the result of compromises or write-offs, which yield the bank little or nothing. NPA reduction is reported under three heads (actual recoveries, "upgradation" or transformation of NPAs into paying assets, and compromises/write-offs). Write–offs involve a complete loss for the banks. According to Finance Ministry figures the share of write-offs in the NPA reduction of the public sector banks rose from an already high 41 per cent in 2014-15 to 46 per cent in 2015-16.

This is also a course for concern since NPAs are often a reflection of wilful default. In the fourth round of what has become a periodic exercise, the All India Bank Employees Association (AIBEA), the "oldest and largest" trade union in the industry, has released a list of 5,610 wilful defaulters on debt they owe commercial banks. The official listing of suit-filed accounts of wilful defaulters disseminated through the Credit Information Bureau (India) Ltd, reports 6,081 cases involving loans totalling Rs. 59,518 crore as of March 31, 2016.

The Reserve Bank of India (RBI) defines a wilful defaulter as one who has diverted bank loans to activities other than the one for which they were originally taken, siphoned funds out with no corresponding assets of any kind to show in the books of the company or who has not repaid loans despite having adequate resources to meet commitments. Thus, the crucial issue here is not default per se, but default that is intentional, deliberate and calculated. This

makes wilful default a criminal offence. A default that results from a wrong investment decision, sheer mismanagement or unexpected changes in the business environment of a firm that does not have the liquidity to meet payments commitments would not fall in the "wilful" category.

Despite this restricted definition, wilful defaults are large. The total default in the 5,610 accounts revealed by the AIBEA adds up to Rs. 58,792 crore. This amounts to around 11 per cent of total non-performing assets in the banking system at the end of March 2016. As many as 4,738 of these accounts accounting for Rs. 47,351 crore of wilfully defaulted loans are with the public sector banks (including the State Bank of India group). In an answer to a Parliamentary question in the Lok Sabha in December 2015, Finance Minister Arun Jaitley reported that in the case of loans of Rs.25 lakh and above from public sector banks alone, the number of cases of wilful default had risen by 44 per cent from 4929 at the end of March 2013 to 7265 at the end of September 2015, and the sum involved by a huge 150 per cent from Rs. 25,804 crore to Rs. 64,335 crore. Clearly, the credit boom during the years after 2004 has been exploited by a set of unscrupulous borrowers, who could avoid scrutiny because of the relaxation in scrutiny that is associated with a post-liberalisation debt spiral.

What is noteworthy about the numbers released by the AIBEA is that the top 106 borrowers (1.9 per cent of the total) responsible for wilful default on loans equal to or exceeding Rs. 100 crore each, together accounted for Rs. 23,093 crore or close to two-fifths (39.3 per cent) of the total sum in default as per this list. Clearly, the NPAs in the books of banks and the manner in which they are being addressed point to a new form of primitive accumulation in India in the Age of Finance.

All of this points to the almost moribund nature of the new capitalism being shaped in India because of increased dependence on foreign finance. In the 1950s, there was near unanimity that winning a degree of autonomy *vis-à-vis* foreign capital was a prerequisite for consolidating India's political freedom. Today, many see recognition by foreign capital as a favoured investment destination as a measure of the country's economic success. But turning moribund and losing sovereignty at the expense of the working people seems to be the real consequence of the accumulated presence of foreign (especially financial) capital in the country since liberalisation.

Author contact: cpchand@gmail.com

SUGGESTED CITATION:
C. P. Chandrasekhar, "Finance capital and the nature of capitalism in India today", *real-world economics review*, issue no. 77, 10 December 2016, pp. 30-39,
http://www.paecon.net/PAEReview/issue77/Chandrasekhar77.pdf

real-world economics review, issue no. 77
subscribe for free

Competitiveness and its leverage in a currency union or how Germany gains from the euro

Thanos Skouras [Athens University of Economics and Business, Greece]

Despite the claims of euro sceptics and the belief of large segments of Germany's public opinion, the dynamism of the German economy has gained considerably from the existence and membership of the euro. The fundamental reason for this gain is to be found in a rather neglected aspect of a currency union's operation, when it is not an optimal currency area. More specifically, when there are considerable differences in the competitiveness of the union's members. The union's exchange rate, which reflects the average competitiveness of the union's members, plays a crucial role in amplifying and sustaining the current account surplus of the members that have higher than average competitiveness. Concurrently, it also amplifies and sustains the current account deficit of the countries with lower than average competitiveness.

It is, of course, widely recognized that the euro's exchange rate has allowed Germany to have a stronger export performance than would have been possible if Germany had a national currency, such as the mark. At the same time, the euro's exchange rate has amplified the current account deficit of the less competitive southern countries. The resulting weakening in their competitiveness, compounded by Germany's reluctance to adopt an expansionary policy, perversely serves today[1] to magnify even further the German economy's gain. The latter benefits by the inflow of capital and labor from the less competitive southern countries, which strengthens the German economy's productive potential while correspondingly weakening further the weaker economies. Thus, the initial gap in competitiveness between the German and the southern economies tends to be increased further.

In order to examine the euro's effect on German competitiveness and, more importantly, to clarify the nature and aspects of the gain, it will be helpful to make a distinction between two different notions of competitiveness.

Two concepts of competitiveness

Before embarking on a discussion of competitiveness, it should be noted that the topic of competitiveness is a controversial one. Despite the widespread use of the notion of competitiveness in public policy discussion and the regular compilation of an international competitiveness index,[2] there is a marked reluctance among academic economists in

[1] Today, the Eurozone is in difficulty because the international financial crisis (originating in the US subprime housing market) made apparent its deficient architecture. In the absence of both a fiscal and a banking union, as well as a credible growth policy, the Eurozone's weakness is likely to persist.
[2] The first well-known index was compiled in collaboration by the Institute for Management Development (IMD) and the World Economic Forum (WEF). Since 1996, the two organizations produce independently their own separate indices; see, WEF (2015) The Global Competitiveness Report 2015-2016, Geneva, World Economic Forum, as well as, IMD World Competitiveness Center (2016) IMD World Competitiveness Yearbook, Lausanne, IMD World Competitiveness Center.

accepting the coherence and legitimacy of the concept.[3]

The reason is that it makes an odd fit with the economic theory of international trade. Ricardo's theory of comparative advantage shows that even if a country is more efficient in producing all goods than another country, it can still gain (indeed both countries can gain) through specialization and trade. The pursuit of competitiveness, in order to outsell the other country and avoid buying from it, does not make any sense. There is no question that the pursuit of competitiveness has a mercantilist provenance and, from Adam Smith onward, the discipline of economics has consistently rejected mercantilism as a norm of economic policy. It is widely accepted among economists that international trade (and, more generally, all voluntary trade) is beneficial to the parties concerned, even though the benefit may be unequally shared.

In fact, contrary to the mercantilist precept, a trade deficit confers a greater immediate tangible benefit than a trade surplus, since the former raises the standard of living and/or investment potential of a country while the latter lowers it. In contrast, the surplus provides claims on future production of uncertain real value. It would seem then that the deficit is clearly preferable. Leaving aside the question of whether a continued deficit is sustainable,[4] it seems beyond dispute that if this were possible it would certainly be preferable.

Nevertheless, a deficit is not clearly preferable to a surplus in a setting of unemployment and spare capacity. In these circumstances, the surplus also increases profits[5] and, through higher profits, encourages production and employment. On the other hand, the deficit reduces profits and tends to lead to lower production and employment. Consequently, in a world of monetary production for profit, often characterized by unemployed resources, a trade surplus makes a lot of sense. This, admittedly, may be a second-best policy but in the real world the first-best policy may, for various reasons, not be feasible.[6] It is in such a context that competitiveness, understood as the ability to consistently achieve surpluses, becomes a useful policy aim.

Competitiveness is a notion, borrowed from microeconomics and competition among firms, that becomes rather complex and hazy when applied to a country. For this reason, a distinction is made below between "essential" and "apparent" competitiveness.

"Essential" competitiveness is, to begin with, analogous to its usage in microeconomics, where it denotes firms' relative ability to compete, and refers mainly to sales cost (including production, finance and marketing costs) but also to other elements, such as product characteristics (including quality, reputation and image), distribution networks, accessibility to

[3] For example, Paul Krugman has argued that "…competitiveness is a meaningless word when applied to national economies. And the obsession with competitiveness is both wrong and dangerous"; see, Krugman, P. (1994) "Competitiveness: A Dangerous Obsession" *Foreign Affairs*, March-April. See, also, Krugman, P. (1996) "Making Sense of the Competitiveness Debate" *Oxford Review of Economic Policy*, Vol.12, No. 3.

[4] Since the deficit is effectively financed by the credit provided by one's trading partners, it is bound sooner or later to provoke a reaction and demand for repayment from the trading partners and creditors.

[5] See, Kalecki M. (1942) "The Determinants of Profits" *Economic Journal*, final version in Kalecki M. (1971) *Selected Essays on the Dynamics of the Capitalist Economy 1933*-1970, Cambridge University Press, Chapter 7.

[6] See, Lipsey, R. G. (2007), "Reflections on the general theory of second best at its golden jubilee" *International Tax and Public Finance*, 14(4), 349-364.

markets and any other factor that contributes to a firm's ability to achieve sustained profitability and do consistently better than its rivals.[7]

In the case of a country, in addition to the above, it includes institutional elements, such as the quality and performance of the education system, the legal and judiciary system, labor relations and the functioning of the labor market, market structure and the degree of monopoly, as well as any other institution that contributes to the country's better economic performance relative to other countries sharing the same currency (or having a long-standing stable exchange rate). Such institutions certainly include the banking and financial system, the health and efficiency of which is crucial to the economy's financing, as well as the efficiency of the state in all its regulatory and other functions.[8]

It is evident that "essential" competitiveness may be affected by changes in any of the above elements. Consequently, it may also be affected by changes in monetary and fiscal policy (as well as other policies and conditions, such as a minimum wage or incomes policy or even prospects regarding political developments), which can have an effect on the level of prices and in the financing conditions and borrowing rates (in the latter case, either directly or through changing perceptions of country risk).

"Essential" competitiveness may be contrasted with the related but somewhat different notion, that of "apparent" competitiveness. "Apparent" competitiveness refers to the ability of a country to compete in international markets with countries that do not share its currency, which depends not only on its "essential" competitiveness but also quite crucially on the exchange rate. Thus, when countries do not share the same currency and exchange rates freely fluctuate (which is the usual condition underlying international trade theory), "apparent" competitiveness not only may be differentiated from "essential" competitiveness but becomes all important.

The comparison between the two notions of competitiveness, with respect to their direction of change, is of particular interest and underlines their difference. "Apparent" competitiveness generally changes in the opposite direction to a change in "essential" competitiveness, when countries do not share the same currency. This is because a change in "essential" competitiveness tends to be compensated by a change in the exchange rate.[9] Thus, for example, an increase in a country's "essential" competitiveness leads to more exports and the ensuing higher demand for the country's currency by foreign importers tends to increase the exchange rate. This makes the country's exports more expensive and its imports cheaper, which is tantamount to a fall in its "apparent" competitiveness. The resulting moderation in exports' attractiveness and concurrent increase in cheaper imports, tends to restore balance in the current account.

[7] The factors, which are relevant to competitiveness, are studied by economists in settings characterized by monopolistic or oligopolistic conditions. Nevertheless, given the strong preoccupation of economists with perfectly competitive markets and the preponderance of the "perfect competition" assumption in most of economic theory, these factors have received more attention in business and especially marketing theory.
[8] The OECD country studies, which examine an economy's macroeconomic conditions and structural aspects, cover in effect all the main elements of "essential" competitiveness.
[9] Given the multitude of factors influencing the exchange rate, especially through capital movements, this is a quite rough tendency and holds on strictly ceteris paribus terms, which are assumed in all attempted comparisons.

In trading within a currency union, the member countries' "essential" competitiveness is evidently of paramount importance in determining the intra-union trade balances. But since members also trade with countries outside the currency union, their "apparent" competitiveness is also important in determining their trade balance *vis-a-vis* their trading partners outside the currency union. In the case of a currency union, unlike the previous case of freely floating exchange rates, an increase in a country's "essential" competitiveness does not lower its "apparent" competitiveness. Within the union, a member country's increase in "essential" competitiveness clearly increases its intra-union export share. As there is no compensating change in "apparent" competitiveness to moderate the effect on exports and to restore balance in the current account, the effect of the increase in "essential" competitiveness is, in fact, leveraged.

Moreover, the increase in "essential" competitiveness is also leveraged in its trading with countries outside the union. This is because the exchange rate of the currency union is determined by the weighted average "essential" competitiveness of all its members, in trading as a bloc with the rest of the world.[10] Unless the particular country's trade is large enough to weigh heavily on the determination of the currency union's exchange rate, the exchange rate is hardly affected and the country's increase in "essential" competitiveness is definitely leveraged. In other words, the compensatory change in the exchange rate, which causes the "apparent" competitiveness to move in the opposite direction of any change in "essential" competitiveness thus moderating its effect, is practically absent when a country is a member of a currency union. This is more likely to be the case for any given change in "essential" competitiveness, the smaller the country's share of the union's external trade and, hence, the smaller its importance in the determination of the union's weighted average "essential" competitiveness (WAEC).

Implications of differences between a country's "essential" competitiveness and the union's WAEC

Before assessing the nature of the gain from competitiveness in a currency union, a comparison needs to be made between a country's "essential" competitiveness and the union's WAEC (which, with the current account over time roughly in balance, tends to be reflected in the union's "apparent" competitiveness). If a country's "essential" competitiveness is higher than the union's WAEC, net exports (i.e., exports minus imports) are given a boost while if it is lower net exports tend to decrease. In other words, a country's "essential" competitiveness that is higher than the union's WAEC, favors the appearance (or enlargement) of a surplus in the country's current account while one that is lower promotes and magnifies a deficit. Thus, in achieving current account balance, the common currency's exchange rate is too low for countries with higher "essential" competitiveness than the union's WAEC and too high for those countries possessing lower "essential" competitiveness than the union's WAEC.

It may be noted that so long as the differences in "essential" competitiveness remain unchanged and the union's current account is in balance, the imbalances in the countries' current account can continue indefinitely. It is also worth noting that such current account

[10] The average competitiveness of the currency union reflects the "essential" competitiveness of all the member countries, weighted by their share of the union's trade with the rest of the world. In reference to the union, it constitutes and may be termed the union's "apparent" competitiveness.

imbalances may exist for all countries. Since it is possible that the union's average competitiveness does not coincide with any one country's "essential" competitiveness, it follows that in such a case the common currency's exchange rate will not be appropriate for balancing any country's current account. The extent to which the exchange rate diverges from the one that is required to balance a country's current account, depends on the extent to which its "essential" competitiveness diverges from the union's WAEC. The wider the range of "essential" competitiveness characterizing the different countries in the union, the greater the possible divergence between the currency's exchange rate and the exchange rate needed to balance a country's current account. In other words, the potential inappropriateness (or "wrongness") of the currency's exchange rate for a country participating in the currency union to achieve balance in its current account, is aggravated, when the differences in "essential" competitiveness among the union members are greater.

It may be surmised that a currency union consisting of countries with markedly different "essential" competitiveness, necessarily results in unbalanced current accounts for at least the countries with the highest and lowest "essential" competitiveness. For these countries, the currency's exchange rate is most likely inappropriate and they will inevitably have a persistent surplus or deficit in their current account. The question is, how does the "wrong" exchange rate affect them?

Let us examine first the case of the high[11] "essential" competitiveness country with a persistent surplus in its current account. The surplus ensures an augmented volume of profits.[12] Thus, *ceteris paribus*, not only profits but also output and employment are all greater (when compared to a situation in which the current account is balanced). In other words, the surplus provides an expansionary impetus to the economy.

Dynamically, it tends to actuate rising wages and prices (especially, the closer capital and labor resources are to full employment). It, therefore, tends concurrently to lower "essential" competitiveness, narrow the gap between "essential" and "apparent" competitiveness and reduce the surplus. The tendency to eliminate the surplus and establish a current account balance, even if attenuated, is nevertheless present. The adjustment mechanism is more sluggish than it would be if the country were not in a currency union but it is still operative. It may even be more sluggish in the case of a country that accounts for a large share of the external trade and weighs heavily in the determination of the exchange rate. In such a case, any given reduction in "essential" competitiveness (due to rising prices) is likely to reduce also the exchange rate, thus slowing down the elimination of the gap between "essential" and "apparent" competitiveness.

Turning now to a country with "essential" competitiveness that is lower than the union's average, the exchange rate is too high for balance in its current account and it will show a deficit. A deficit implies smaller profits and, ceteris paribus, smaller profits imply lower output and employment. Thus, the deficit imparts a contractionary bias to the economy. The dynamic tendency is for lower wages and prices to reduce the deficit, improve the "essential" competitiveness and, by bringing the latter closer to the union's average, promote adjustment. The adjustment mechanism is blunter than it would be in the absence of the currency union but it is still in operation.[13]

[11] High relative to the union's average.
[12] See, Kalecki, *ibid*., for the conditions under which profits increase by exactly as much as the surplus.
[13] It is as if a speed bump slows down the road to adjustment.

Despite the existence of an automatic adjustment mechanism, a currency union with a balanced foreign account can shield members from correction by market forces, if they choose to override this mechanism. By following policies that lead to high or low "essential" competitiveness, it allows surpluses (by the highly competitive countries) and deficits (by the least competitive countries) to develop and be sustained without hindrance. Such imbalances may in fact continue practically indefinitely, so long as the associated intra-union capital flows are not blocked and allowed to go on.

The nature of Germany's gain from the euro

In order to elucidate the nature of Germany's gain from the euro, a brief historical detour needs to be made.

It is widely believed that the euro was not only an economic project but also, if not primarily, a political one that was initiated by France rather than Germany.[14] The latter assented to the French demand for a decisive step in European integration, as a show of good will and a price to pay for German re-unification. The process of re-unification proved to be quite a burden to Germany and, at the time of the euro's introduction about a decade later, its "essential" competitiveness was at a relatively low ebb, its growth rate lagged the European Union average, its current account was in deficit and it was considered "the sick man of Europe".[15]

It was only after the introduction of the euro and its initial sizeable fall against the US dollar,[16] that Germany's current account was balanced (following a decade of being in deficit), while its growth performance remained weak in the first years following the introduction of the euro. The turning point seems to have taken place in 2004. From that year onward, Germany's economic performance began to improve decisively, becoming clearly and increasingly superior to the Eurozone's average.

It is still debated whether this turnaround is primarily due to the immediately preceding Hartz reforms, which deregulated the labor market with a series of measures between 2002 and 2005, or whether it was the result of longer-run forces, which gradually transformed the German economy. The Hartz labor market reforms reduced the duration and level of the unemployment benefits and established the institutional preconditions for the enlargement of part-time employment and the creation of a segmented, dual labor market. This probably contributed to the containment of wage rises and to a lesser extent of unit labor costs, relative to the Eurozone average, but it has been disputed whether it has been the main influence in the determination of Germany's economic performance.[17] It seems more likely that the

[14] For an official recognition, see Van Rompuy, H. (2014) *Europe in the Storm: Promise and Prejudice*, chapter 2, Davidsfond, Uitgeverij.
[15] See, The *Economist* (1999) "The Sick Man of the Euro", June 3. But even then, Germany's "essential" competitiveness was not lower than the European average.
[16] From its introduction on 1/1/1999 at a value of $1.179, the euro dropped to a minimum of $0.825 on 25/10/2000 before gradually recovering its initial value by May 2003. (It reached the highest ever value of $1.60 in July 2008, fell under $1.10 in early 2015 and is hovering in the $1.10-1.14 range since then).
[17] See, Storm, S. and C.W.M. Naastepad (2015) "Germany's Recovery from Crisis: The Real Lessons" *Structural Change and Economic Dynamics*, 32 (1): 11-24. Also, Storm, S. and C.W.M. Naastepad (2015) "Myths, Mix-ups and Mishandlings: Understanding the Eurozone Crisis" at: http://ineteconomics.org/ideas-papers/research-papers/myths-mix-upsandmishandlings-what-caused-the- eurozone-crisis These economists argue convincingly that labor unit cost was of little import to the development of a surplus in the balance of payments and that superior technology and high quality of industrial products, in combination with the rapidly rising demand for such products from China and

process of transforming the economy started much earlier, in the early nineties if not before, with the gradual internationalization of the German industry. The shifting of production abroad initiated the containment of unit costs and the international outlook of the German economy.[18] Moreover, the flexibility of industrial relations, which permit collective bargaining at the level of the firm, and the traditional institutional characteristics of the German labor market seem to be more important than the Hartz reforms, in explaining the performance of the German economy.[19]

Nevertheless, the Hartz reforms have acquired an important symbolic role, in demonstrating the German political elite's determination to safeguard and promote competitiveness in the context of the Eurozone. It must be remembered that they were instituted by a socialist government, which early in its term resisted austerity and committed the first ever transgression of the Maastricht Treaty, by running repeatedly a budget deficit exceeding 3% of GDP. This same government proved willing to confront the trade unions over the unpopular Hartz reforms, which in all likelihood contributed eventually to its electoral defeat. Thus, whether the reforms actually improved or not the economy's performance in a crucial way, a strong message was given that competitiveness is the paramount objective for the whole political class, including the socialists.

It may be noted that the strong concern with competitiveness was not a new development but a continuation of a tradition which dates at least since the first years after the Second World War.[20] Throughout the post-war period, Germany's economic policy was steadfastly orientated to the achievement of surpluses in the current account. Consequently, competitiveness has been traditionally a strong concern and the institutional elements of competitiveness were well developed at the time of the Eurozone's creation. Moreover, the still fresh in memory experience of re-unification and consequent collapse of the East German economy following the adoption of a common currency, demonstrated quite clearly the crucial importance of competitiveness in a currency union.[21] Thus, the socialist government's determined implementation of the Hartz reforms confirmed beyond any doubt that competitiveness would continue to be a prime concern in the new Eurozone era. In these circumstances, given the German economy's quite high "essential" competitiveness at the Eurozone's start and favorable conjuncture of a strong demand from China and other emerging markets for Germany's technologically advanced machinery products, it is not surprising that exports reached 50% of GDP.

---

other growing emerging markets, were much more important factors in explaining developments in the balance of payments. They also point to the differences in productivity between Germany and the Eurozone, which have been increasing since the re-unification of Germany at least a decade before the Hartz reforms.

[18] For an informative analysis of the German political economy, see Barasin, C. (2013) *Germany: A Global Miracle and a European Challenge*, Brookings Global Economy and Development, Working Paper 62, May.

[19] See, Dustmann, C., Fitzenberger, B., Schönberg, U., & Spitz-Oener, A. (2014) "From sick man of Europe to economic superstar: Germany's resurgent economy" *The Journal of Economic Perspectives*, 28(1), 167-188.

[20] From 1950, if not earlier, Ludwig Erhard emphasized the need for the "internal discipline" that could ensure price stability, in order to achieve the prime target of export growth. See, Sauramo P. "Germany's Success: A Finnish Perspective" in Unger B. ed. (2015) *The German Model: Seen by its Neighbours*, SE Publishing.

[21] Chancellor Kohl's insistence on the exchangeability of the Deutschemark and the East German mark at a rate of one to one was the final nail to the coffin of the East German economy. This parity "did not correspond to reality", in the view of respected central banker Karl Otto Pohl and seems to have provoked his resignation from the Bundesbank in May 1991, mid-way through his second, eight-year term.

In conclusion to this historical detour, Germany entered the Eurozone with a head start in "essential" competitiveness and begun early on a politically arduous, sustained effort (from 2002 to 2005) to implement structural reforms in order to further augment it. Thus, Germany benefited from the start and, whether as a result of this effort to reform the labor market or, more likely, the international conjuncture and its prior industrial specialization, it managed to transform itself from the "sick man of Europe" to the "model economy of Europe". Germany's claim to having become a model economy is based on the fact that in 2007 it achieved a current account surplus equal to 7% of GDP and managed to overcome the 2008 crisis first among all European economies, surpassing the pre-crisis GDP in just three years. This success is closely linked and, to a large extent, due to Germany's participation in the Eurozone and, more specifically, to the leverage of Germany's competitiveness made possible by the Eurozone.

Membership in the Eurozone tends to reinforce the "essential" competitiveness of a country (like Germany) that has higher "essential" competitiveness than the union's average and to lessen the "essential" competitiveness of a country (such as Greece) that has lower "essential" competitiveness than the union's average. To put it differently, and more generally, membership enables those countries which are more "essentially" competitive than the average to have an exchange rate that is lower than would have been the case if they were not members, thus leveraging their "essential" competitiveness vis a vis other countries (including, possibly most significantly, the rest of the world).[22] Conversely, membership obliges countries with below the Eurozone's average "essential" competitiveness to have a higher exchange rate than what would be the case outside the currency union, thus reducing further their "essential" competitiveness vis a vis all other countries.[23]

An analogy from the world of competitive sports may be helpful here. The handicap system used in diverse sports, such as golf or horse races, is analogous to the movements in the exchange rate following a change in relative "essential" competitiveness. For example, according to horses' relative performance in training and in previous races, they are saddled with different weights, so as to equalize the chances across all horses competing in a race. In this setting, a currency union is like grouping the horses of a stable together and assigning a single common handicap to all of them, on the basis of their average performance. Such an arrangement, would obviously grant the best performing horses of the stable an unfair advantage and, by the same token, disadvantage the worst performing ones when competing with other horses, which have been given a handicap based on their individual performance.

Given that the Eurozone's current account has been roughly in balance (or, more recently, small surplus), the euro's exchange rate tends to roughly reflect the Eurozone's overall competitiveness relative to the rest of the world. The euro's exchange rate has resulted in large surpluses for Germany and, to a lesser extent, some Northern European countries and equivalent deficits for mostly Southern European countries. In other words, the euro's exchange rate was too low for balance in the current account of Germany and the North, while being too high for balance in the current account of the South. In this way, the union's exchange rate has leveraged in opposite directions the "essential" competitiveness of member countries; it augmented the "essential" competitiveness of Germany and the North and weakened the "essential" competitiveness of the Southern countries.

[22] It is not *all* other countries because countries that are members of the union and happen to have higher "essential" competitiveness gain an even bigger leverage, since their exchange rate would have exceeded the union's exchange rate by a greater margin.
[23] Except those members of the union, which have an even lower "essential" competitiveness.

This effect becomes evident if we imagine Germany outside the Eurozone. Then, its exchange rate against the US dollar would tend to be higher and, as a consequence, it would be "effectively" less competitive. More importantly, unlike what has been the case whilst a member of the Eurozone, Germany would find that a surplus in its foreign account is not sustainable. Any attempt to manipulate the exchange rate and fix it at a lower level, would provoke protest and retaliation from its trading partners. Conversely, for the remaining countries within the Eurozone, the euro's lower exchange rate would increase their "apparent" competitiveness and their current account would be improved. As a result, the increase in aggregate demand from higher net exports would enable their economies to grow.

Such a course of action for reviving the Eurozone economies has, in fact, been proposed by George Soros, who has argued repeatedly since 2012, that if Germany is not willing to bear the cost of leadership in the creation of a federal European state, it should leave the Eurozone.[24] This is contrary not only to the economic but also to the political interest of Germany, which has managed through the Eurozone to become the undisputed leading power in Europe.[25] Moreover, the Euro crisis has provided in a number of ways an additional economic benefit to Germany.

To start with, the contractionary policies and high unemployment in the Southern countries, and particularly in Greece, have caused a sizeable influx of young immigrants from these countries into Germany. This alleviated the tightness in the German labor market and, given that a high proportion of these immigrants (especially from Greece[26]) were highly educated and well trained, this addition to the German labor force represented a considerable economic gift from the Southern countries (which had borne the cost of raising and training this labor force) to Germany.

A second major way in which Germany benefited from the Euro crisis, was from the inflow of capital from the Southern countries. In particular, the fear of Grexit and its potential domino effect have prompted a large transfer of capital from the South, in the direction of German government bonds, the Frankfurt stock exchange and real estate in various German cities. This had a stimulating effect on the German economy and resulted in an increase in the wealth of German equity and real estate owners, as well as a reduction in the cost of finance for both German firms and especially the German state.

An estimate of the benefit of lower borrowing costs for the German state, that was occasioned by the Greek crisis, indicates that it easily exceeds the cost of the Greek crisis to the German taxpayers.[27] More specifically, the study estimates how the German bunds' price reacted to disquieting news concerning developments in Grexit prospects between the spring of 2010 and the summer of 2015. Such news increased the demand for and the price of bunds, thereby lowering interest rates and borrowing costs. The benefit to the German state from news worsening and increasing insecurity is estimated at well over 100 billion euros, while the total commitment through all channels (including the European Stability Mechanism) of the German state to the rescue of the Greek economy was about 90 billion euros until the middle of 2015. This benefit represents about 3% of German GDP and certainly contributed to the balancing in 2014, for the first time in many decades, of the German government budget.

---

[24] See, *Spiegel Online*, 26 June 2012, *www.spiegel.de English Site › Europe › Euro Crisis*
[25] See, Beck U. (2013) *German Europe*, Polity Press.
[26] It is estimated that about 3,500 medical doctors and 30,000 scientists have emigrated from Greece in the last 5 years; see, www.huffingtonpost.gr/loislabrianidis/_2408_b_8520596.html
[27] Dany G., Gropp R. E. and von Schweinitz G. (2015), Germany's Benefit from the Greek Crisis, Leibniz-Institut für Wirtschaftsforschung Halle, IWH Online 7/2015. Halle (Saale).

The total benefit to the German economy, including the gain for German firms from lower labor and borrowing costs and the increased wealth of equity and property owners, is not easy to estimate but is certainly considerably greater. Of course, it is not evenly shared and this probably accounts for the fact that it is not recognized by the German public opinion. Those who evidently gain are the equity owners, real estate property owners and employees of large export-oriented firms; the gain for the rest of the working class and especially the lowly-paid part-timers is disputable while interest earners, such as savers and insurance organizations, seem to lose out.[28]

A third source of benefit emanating from both the euro and Greek crises, is the euro's weakness, which has resulted recently in a current account surplus of 3% of the Eurozone's GDP. The euro's weakness increases the "apparent" competitiveness of all Eurozone countries and leverages further Germany's "essential" competitiveness. The benefit from the weaker euro is more pronounced for those countries, such as Germany, that trade the most with the rest of the world.

Finally, an implication of the above is that Germany benefits also on a broader important issue of political economy. This is because uncertainty, following the shattering of unfounded confidence by the euro crisis, is inherent to the present state of the Eurozone and can be overcome only by a determined advance in the completion of banking and fiscal union and a decisive step towards federal Europe. A protracted state of uncertainty surrounding the Eurozone's prospects is damaging to the weaker members but not to Germany. Consequently, Germany has an invaluable bargaining advantage in determining the shape of any federal solution; in contrast to the damage incurred by other members, stalling until it gets its own way is for Germany not only costless but actually beneficial.

To briefly recapitulate and conclude, Germany gains considerably from its membership of the Eurozone, through leveraging of its "essential" competitiveness. In addition, it further gains from the euro crisis and the continued uncertainty about the Eurozone's future. These gains can be maintained, so long as the Eurozone is not totally dismantled but continues in some form with Germany being a member. All that is required for Germany to be able to gain from the leverage, is that its "essential" competitiveness remains higher than the Eurozone's weighted average. Of course, the greater the difference between its "essential" competitiveness and the Eurozone's average, the greater the leverage and the gain. Finally, though the uncertainty over Grexit served well the purpose so far, it is not essential for gaining from uncertainty. The preservation of uncertainty by some other means may also be effective, so long as Germany retains its superior "essential" competitiveness.

Author contact: savthan@gmail.com

SUGGESTED CITATION:
Thanos Skouras, "Competitiveness and its leverage in a currency union or how Germany gains from the euro", *real-world economics review*, issue no. 77, 10 December 2016, pp. 40-49,
http://www.paecon.net/PAEReview/issue77/Skouras77.pdf

[28] The loss of insurance organizations from lower interest needs to be set against their gain from the increase in contributions paid by the greater number of workers (due to both lower unemployment and more immigrants).

real-world economics review, issue no. 77
subscribe for free

Rethinking Rationality Theory's epistemological status. Normative vs Positive approaches

Gustavo Marqués & Diego Weisman [FCE–UBA, CIECE–IIEPBaires; CIECE–FCE, UBA; UNLZ]

1. Introduction

In 1944 John von Neumann and Oskar Morgenstern proposed an axiomatic theory of choice under risk – known as Expected Utility Theory (EUT) – which implied the Bernoullian expected utility hypothesis. Their proposal attracted many economists because it solved the apparently unmanageable problem of how to measure utility and, under well specified conditions, allowed predictions about the choices that an individual would make among alternative lotteries. In the past utility theory had been successfully applied to economics in situations in which certainty prevailed (like consumption), but it was considered useless under conditions of risk and uncertainty. Neumann – Morgenstern's theory, as well as other equivalent theories developed shortly after its publication, seemed to open the door for the formidable task of unifying the field of economics. Hopefully, adequately developed it could help to explain decisions involving risk like gambling and insurance, which at that time were considered beyond the scope of economic theory and originated in arbitrary behavior or the prevalence of psychological factors.

A few years after EUT was proposed a debate about its epistemic status started. Some authors, like Milton Friedman, assumed the positive-normative distinction, traditional in economics, and considered that EUT was merely an empirical (descriptive) hypothesis.[1] Other economists (among them Savage (1954), Marschak (1950), Strotz (1953) Ellsberg (1961) and Raiffa (1961)) found the axioms of EUT persuasive and plausible at the normative level. If they were right, economics might count at last with a *general* theory of *rational* decision, applicable to certain as well as risky and uncertain choices.[2] All of them took for granted that the descriptive and normative adequacy of EUT were not only desirable but also perfectly compatible and reachable goals. In the following sections the importance of counting with a well confirmed rational theory of decision making will be examined and its feasibility critically revised. We also point out that the normative dimension of EUT was misunderstood by leading economic figures, and the crucial distinction between paradoxes and anomalies overlooked.

[1] See Friedman and Savage (1948, 1952), where Friedman's methodological emphasis clearly prevails. Friedman's most complete account of his epistemic vision of economic theories may be found in his 1953 article.
[2] "These axioms have strong intuitive appeal. It would seem that every normal person would clearly accept them as precepts of behaviour" (Strotz, 1953, p. 391). "...Savage's theory is not a descriptive or predictive theory of behavior. It is a theory which purports to advise any one of its believers how he *should* behave in complicated situations, provided he can make choices in a coherent manner in relatively simple, uncomplicated situations" Raiffa (1961, p. 690).

2. Normative analysis of EUT. *Feasibility* and *adequacy* of a norm

Though Bernoulli had made some indications about the normative value of his Expected Utility Hypothesis, he regarded it mainly as descriptive in nature. Later decision theorists, like Neumann and Morgenstern themselves, and, even more explicitly, Savage (1954), Marschack (1950) and Strotz (1953), underscored more explicitly its normative import. According to this view, EUT is not only a descriptive theory about how people make *in fact* decisions, but also a prescriptive set of rules about how decisions *should* be rationally made.

> "Consider any person not deemed insane who holds contradictory preferences…. Imagine that we explain to this person the nature of the contradiction, pointing out clearly how his preferences violate our axioms. Will he in consequence of understanding the nature of the contradiction decide that his preferences are ill-founded and proceed to change them, or will he persist in his original preferences even though it is entirely clear to him exactly what precepts his preferences violate. If for nearly every person holding contradictory preferences an understanding of the character of the contradiction induces him to straighten out his preferences, then the Neumann-Morgenstern axioms may properly be regarded as precepts of rational choice. My own feeling is that it would be a strange man indeed who would persist in violating these precepts once he understood clearly in what way he was violating them" (Strotz, 1953, p. 393).

Indeed, both properties – the *normative* as well as the *descriptive* character of EUT – were considered not only compatible, but also intimately related. The hopes of Savage, Marschack and Strotz regarding its empirical adequacy were probably based on what they believed was its normative value: regularly "normal" individuals can learn from their mistakes. As Strotz (1953, p. 393) said, "my own feeling is that it would be a strange man indeed who would persist in violating these precepts once he understood clearly in what way he was violating them". As long as Expected Utility Theory injects rationality upon people's behavior, eventual deviations from the correct choices could be gradually eliminated. Though EUT might be initially inadequate from a descriptive point of view, its influence upon agents' decisions would result in a progressive adjustment between its prescriptions and the observed choices of individuals. Its normative adequacy guarantees its self validation at the empirical level.

Is this happy dream of a EUT sustainable on empirical and normative grounds likely to come true? To be satisfactory any prescription should comply with two main requisites: a) *feasibility* (individuals must be able to follow the instructions received from the theory); b) *adequacy* (its instructions must be "reasonable" or unassailable after rational analysis). Both conditions are necessary for the normative adequacy of any axiom of EUT. But they are independent: granted the feasibility of EUT's axioms their adequacy must still be proven. Below, we examine first the intuitive plausibility of EUT's axioms and later the legitimacy of its consequences.

2.1. Normative plausibility of axioms and assumptions of EUT

How good is EUT as a theory of *rational* choice?[3] The theory has not only axioms (explicit postulates) but it also involves assumptions. It is not totally clear what the logical status of an assumption is but if the axioms and assumptions of EUT have been identified, as some authors claim, in some cases its normative assessment may be carried out *directly* by employing the two criteria described above. A useful and comprehensive description of the content of EUT is provided by Starmer (2000, 2 and 3), and his classification might be sketched as follows:

Axioms	**Assumptions**
1) Completeness	1) Procedure Invariance
2) Transitivity	2) Description Invariance
3) Continuity	3) Stochastic Dominance
4) Independence	

It is beyond our purpose to undertake a thorough analysis of every postulate of EUT, but a brief indication of how to use these outlined criteria would be helpful. Transitivity, for instance, seems *feasible*, in the sense that no obstacle prevents individuals to understand the rule and follow it consciously. It is also easy to defend its *adequacy* resorting to the money pump argument. In the case of assumptions, stochastic dominance seems also unassailable: as soon as individuals are aware that option A dominates option B they choose the first alternative. If every action and assumption of EUT could be defended in this way its legitimacy would be beyond question.

There are cases, however, where those criteria cannot be applied. Continuity and Procedure Invariance, for instance, cannot be easily understood as rules or precepts for rational action and are best defended as technical conditions that should be fulfilled for EUT's mathematical tractability. Unfortunately, a sharp demarcation among precepts of rational conduct and technical requirements cannot be formulated in logically acceptable terms: both are indistinctly treated as necessary conditions for the attribution of a utility function to individual agents. Logic alone is unable to show the difference between the two types of sentences.

2.2. Baumol and Ellbergs' objections to EUT and SEUT

There is another way, this time *in*direct, of assessing the normative adequacy of EUT, which puts its *consequences* at the center of the analysis. In his 1951 article Baumol objected to the current efforts for pushing forward the scope of the expected utility hypothesis arguing that in so doing excessive restrictions on preferences were incorporated.

> "The point is simply that the assumptions of the system, in Samuelson's happy phrase, 'put a straitjacket on the person's preferences'. Once he has made up his mind on a few things, the rest is decided by him for the rules. From his choices among some limited sets of items we know how he will react to the rest, and there is

[3] This question would have been interpreted as rhetoric for most working economists after the Neumann-Morgenstern revolution. In allusion to SEUT, Ellsberg (1961, p. 650) pointed out that "In general, as one ponders these postulates and tests them introspectively in a variety of hypothetical situations, they do indeed appear plausible. That is to say that they do seem to have wide validity as normative criteria (for me, as well as for Savage)".

> no apparent reason why choice should be so circumscribed in fact" (Baumol, 1951, p. 64)

In fact, EUT allows a decision maker to choose freely among a few initial options, but having done this, his following elections are constrained by the previous ones. These restrictions, argues Baumol, are inconvenient because "it is not at all difficult to construct examples in which the Neumann-Morgenstern index leads to results conflicting with plausible preferences" (Baumol, 1951, p. 64). Consider for instance a decision maker who has to choose at first between three lotteries A, B and C whose expected utilities are, respectively, 600, 420 and 60 "utils". In a second round he is asked to make another election among the two following lotteries:

a) (A, 5/6; C, 1/6) b) (A, 1/6; B, 5/6)

If the subject were a believer in EUT he would find that U(a) = 510 "utils" and U(b) = 450 "utils", and consequently he would pick a over b. "Yet who is to say that preferring b to a is pathological? Is it unthinkable that some, or even many, will prefer the insurance value of having 420-util situation in the hand to the 600 in the shrubbery, especially since it is the totality of the individual's possessions which is to be decided on?" (p. 64). Incidentally, it is interesting to note that Baumol is referring here to the "segregation" heuristic discovered later by Kahneman and Tversky as an essential part of the so-called process of Editing.

Adopting a similar stance, Ellsberg (1961) imagined a set of hypothetical choices, this time under conditions of uncertainty or, as he labeled them, *ambiguity*. This takes us to a further development of the theory known as Subjective Expected Utility Theory (SEUT). In one of the cases described by Ellsberg the individual had to pick up a ball from a box which contained 30 red balls and another 60 balls which were either black or yellow in unknown proportions. The subject received $100 in case he extracted a ball that had the color on which he had bet and nothing otherwise. The four options faced by the decision maker are represented below:

	30	**60**	
	Red	**Black**	**Yellow**
I	$100	$0	$0
II	$0	$100	$0
III	$100	$0	$100
IV	$0	$100	$100

In this thought experiment the subject had to choose first between alternatives I or II, and then between III or IV. Choosing I means that he bets (believes) that a red ball will be selected but if he chooses II his bet implies that he believes the extracted ball will be black. Going to the following stage, choosing III means that he is betting for "red or yellow" and selecting IV for "black or yellow".

Ellsberg observed that the more frequent pattern consisted in preferring I to II at first, and IV to III in the second stage, which violates the "sure thing principle" of Savage's axioms. Any individual who chooses in this way behaves inconsistently. If he bets on I in the first stage, it means that he believes that the probability of extracting a black ball could not be greater than

1/3 (otherwise he would have bet on II). But then he should have bet for III, not IV in the second stage, because he should have believed that the probability for the chosen ball to be black or yellow could not be greater than the probability that it was red or yellow. Ellsberg concludes that in cases like these choices do not reveal probabilities, as Savage pretended. In his words:

> "for any values of the pay-offs, it is impossible to find probability numbers in terms of which these choices could be described – even roughly or approximately – as maximizing the mathematical expectation of utility" (p. 655).

Until now, supposing that the "experiment" has been correctly devised, Ellsberg has found out an *anomaly* of SEUT. This finding has little meaning in itself, because the deviation may be caused by irrational behavior on the part of individuals. But suppose the agents are shown their inconsistency and are then allowed to reconsider their choices in light of the new information. Taking as granted the *normative* adequacy of SEUT, they should modify their pattern of choices in a way consistent with Savage's axioms. However, many decision-makers *do not rectify* their previous decisions even after knowing they were mistaken according to the theory.

> "The important finding is that, after rethinking all their 'offending' decisions in the light of the axioms, a number of people who are not only sophisticated but reasonable decide that they wish to persist in their choices. This includes people who previously felt a 'first order commitment' to the axioms, many of them surprised and some dismayed to find that they wished, in these situations, to violate the Sure-Thing Principle. Since this group included L.J.Savage, when last tested by me (I have been reluctant to try him again), it seems to deserve respectful consideration" (Ellsberg, 1961, p. 656).

The important point in Baumol and Ellsberg's "experiments" is that the apparently reasonableness of the "biased" choices transforms the anomalies into *paradoxes*. The so-called paradoxes of EUT or SEUT are relevant for its *normative* evaluation because in these cases (unlike what happens with the presence of anomalies) the elections made by the individuals do not appear to be irrational. The presence of a paradox does not mean that the subjects cannot or do not whish to adjust their choices to the axioms of the theory; it rather indicates that even if the individuals had the super-human capacities needed for behaving as the theory instructs them, some of the results reached in so doing would be unacceptable anyway. The problem with paradoxes is not that as *a matter of fact* the subjects "deviate" themselves from the standards. The problem arises because the "biased" observed behavior may be considered at least as rational as the one prescribed by the theory. What is at stake then is the *adequacy* of EUT and SEUT.

The presence of paradoxes is important in another sense: they indicate that two different notions of rationality interfere with each other: the one defined by the theory, which disqualifies as irrational any deviation of the standards posed by itself, and another – intuitive– notion of rationality which allows a defense of some choices considered biased from the standpoint of EUT.[4] So, anytime a paradox arises it carries with it two main implications:

[4] Ellsberg (1961, p. 656) probably had a similar idea in mind. Confronted with paradoxes he considered that "the choices themselves do not appear to be careless or random. They are persistent, reportedly

a) an intuitive theory of rationality inconsistent with the traditional one; b) in this particular case such a theory performs at least as good as (and sometimes better than) the more formal theory incorporated in EUT.

Let's look more closely at the relation between these two concepts of rationality. Initially, given any definition X, one could decide whether a particular entity belongs or not to its domain. As long as EUT is empowered for defining "rationality", any singular choice should be assessed by its standards. Any time a definition is used it is posed beyond discussion: its application assumes its validity. How could a definition be put in question? One way is by first identifying – in an intuitive manner – a number of admittedly rational choices and look then whether they belong to the domain of the definition. This procedure is the opposite to the one mentioned before, where the definition constituted, so to speak, its own domain. In this second case, instead, the domain of rationality is set first (at least partially) and later a search for a definition that conforms to it begins.

Usually, however, an intermediate way is followed which allows for a feed-back process between the definition and its intended domain. The sequence could be as follows. First a vague and intuitive idea of rationality is advanced and with its help some behaviors are identified as rational. Second, a more formal definition of rationality (for instance, SEUT) is designed with the purpose of increasing the accuracy and precision of the former notion. This procedure is called *elucidation*. A good elucidation should preserve in its domain most of the behaviors pertaining to the domain of the vague notion, but not necessarily all of them, because the new concept corrects the older one in some respects. The domain of the elucidating concept is not necessarily co-extensive with the one of the elucidated notion. Besides, it is expected from a good elucidation that in cases in which maladjustments with the older notion are found, further inspection will reveal its superiority.

One can say that these two notions of rationality share their "power". The elucidation restricts the scope of the primitive notion, but at the same time the latter retains its power imposing restrictions and limits on the formal notion. Which one is more powerful is a relative point. Once a formal definition has been imposed people apply it without judging its merits, so that it works as a benchmark for rationality. But when the construction of the formal notion has taken place the benchmark is provided by the vague and intuitive notion. Some of this can be observed in Plato's dialogues where Socrates asks his occasional interlocutor to advance some definition (for instance, of courage or goodness) showing later that it is ill conceived because it either leaves some important phenomena outside its intended domain or includes some others which shouldn't pertain to it. In all these cases the intuitive domain works as a guide for further modifications of the more formal definition.

A normative assessment of EUT cannot be founded merely on bare empirical evidence. To assess the merits of any rational theory of decision making another normative notion for choice-making is required as a benchmark of rationality. Taken by itself the empirical evidence is irrelevant for normative purposes and should be interpreted normatively for playing such a role. These considerations throw light on the important differences among

---

deliberate, and they seem to predominate empirically; many of the people who take them are eminently reasonable, and they insist that they *want* to behave this way, even though they may be generally respectful of the Savage axioms. These are strong indications, in other words, not merely of the existence of reliable patterns of blind behavior but of the operation of definite normative criteria, different from and conflicting with the familiar ones, to which these people try to conform".

anomalies and paradoxes. Besides, their impact on the normative value of EUT is unequal. Anomalies could gradually disappear as long as people learn to take decisions correctly, but this fact will leave the value of paradoxes untouched. The traditional post Neumann-Morgensten's project aimed at the construction of a theory of *rational* decision making will be more endangered by the presence of paradoxes than that of anomalies.

3. EUT as a merely descriptive theory

In two papers written shortly after the Neumann-Morgenstern contribution, Friedman and Savage took as given the positive-normative distinction between sentences or theories disregarding altogether the *normative* dimension of the axioms of EUT. They distinguished also between EUT and what they called Expected Utility Hypothesis (EUH), and focused on the positive problem of the descriptive adequacy of EUH.[5] In so doing they defended a view of the epistemic nature of Expected Utility Hypothesis which greatly differs from the one that was examined before. There are two different interpretations of the positive approach towards EUH: the first one assumes that it describes the rule that the agents as a matter of fact follow when taking decisions in non-certain settings; according to the second interpretation it describes the objective results of the individual choices (at the individual and aggregate levels). Friedman and Savage endorse this last position:

> "The hypothesis does not assert that individuals explicitly or consciously calculate and compare expected utilities. Indeed it is not at all clear what such an assertion would mean or how it could be tested. The hypothesis asserts rather that, in making a particular class of decisions, individuals behave as if they calculated and compared expected utility and as if they knew the odds" (F-S, 1948, p. 298).

Leaving out of consideration the normative dimension of Neumann-Morgenstern's theory had a price for Friedman and Savage: they fell short of fully appreciating the crucial point in the criticism of Baumol referred above. Their reply was indeed disappointing. They got rid of Baumol's position by pointing out that to imagine a possible counterexample for the theory means it is not void of empirical content and blamed Baumol for seeking an unfalsifiable theory of decision making under risk[6]. It is true that Baumol had confusing ideas about the methodological importance of having empirical content, but one thing is to clarify this issue and another thing is to misrepresent the meaning of Baumol's counterexample.

The imaginary objection of Baumol is not a simple (conceivable) anomaly that might arise within EUT's frame. In Popperian terms, taken in great consideration at that time, its counterexample is not a potential falsifier of EUT but rather a *normative* falsifier: it shows that EUT may have consequences inconsistent with patterns of choice which seem to be nonetheless perfectly reasonable. This finding put into question the standard notion of

[5] EUT is logically equivalent to EUH, but this last expression makes explicit a concrete operation for calculating the expected utilities of lotteries. As usual EUH may be represented as $\sum$ pi u(xi).
[6] Friedman and Savage (1952, pp. 465-466) make a distinction between conceivable behavior that would contradict the hypothesis and behavior that has been observed and contradicts the hypothesis. Referring to the first type of "evidence" they say that "this feature is clearly a virtue of a scientific hypothesis, not a defect – it is a valid objection only to claim that the hypothesis must be true (i.e., is a tautology). The possibility of specifying behavior that would contradict the hypothesis means that the hypothesis is not empty".

rational choice under risk. Friedman and Savage failed to notice the difference between a simple conceivable *anomaly* (which is a virtue of any theory, at least in the minimal sense that is not completely void) and a *paradox* (which, even if just conceivable and not yet realized, contests the axioms of EUT as precepts of rational action). They didn't see the difference, we guess, because they intentionally pushed the normative approach of theories beyond their consideration[7].

Interestingly, their neglect *weakened* the content of EUT. A whole set of paradoxical choices which are decisive for the assessment of the theory as long as normative adequacy is required, are dismissed as beside the point if a merely descriptive point of view about theories is subscribed. Friedman and Savage could say then that Baumol's objection only shows that "the expected utility hypothesis is *potentially* fruitful" and, consequently, his critic "is largely irrelevant".[8]

4. EUT as a *descriptive* theory of *rational* choice

During the second half of the twentieth century as well as in recent years a good number of experiments designed to test EUT and SEUT have shown the recurrent presence of anomalies in peoples' behavior. These findings, especially those obtained by Kahneman and Tversky, have put into question the golden dream of a *rational* theory of decision making satisfactory from a descriptive point of view. Whatever the philosophical or epistemological position one can endorse regarding the axioms or assumptions of the standard theory, anomalies are undesired and must be eliminated. Four strategies for dismissing as irrelevant the anomalous findings have been suggested.

a) It is claimed that the so-called anomalies have been obtained in ill-designed circumstances characterized by insufficient incentives. Supposedly they will tend to disappear as soon as adequate incentives are restored. However, this trend has not been observed in many important cases. In a well-known experiment, Lichtenstein and Slovic (1971) found preference reversals using small amounts of money as compensation. When Grether and Plot (1979) replicated their experiment using this time substantive monetary incentives the phenomenon not only did not disappear but was accentuated. Though it is true that in some cases the use of incentives helps to reduce the biases, the general thesis that they always work in this way (and, especially, the idea that proper incentives will in the end completely eliminate the anomalous phenomena) is yet an unconfirmed hypothesis.

[7] "Although the word 'paradox' is frequently used in the history of EUT, it should be clear from this and the previous examples that it does not refer to deeply engrained conceptual difficulties, such as Russel's paradox in set theory, or the EPR paradox in physics, but rather just to problems or anomalies for the theory that is currently taken for granted" (Mongin, 1998, p. 172). Mongin is correct in that "paradox" means something different in logic and in economics, but he is wrong in saying that the paradoxes of EUT are just anomalies. The Friedman-Savage's neglect is still alive.

[8] Friedman and Savage, 1952, p. 466. It is worth noticing that even if the descriptive approach of Friedman-Savage is accepted, and paradoxes are treated as mere anomalies, they are still relevant because if confirmed they show that a good number of people do *not* behave *as if* they were trying to maximize their expected utility. The practice of reducing paradoxes to anomalies does not suffice for putting the Friedman-Savage point of view beyond criticism: anomalies (which defy the descriptive credentials of the theory) are far more dangerous in their approach than paradoxes (which questions its normative value).

> "It has frequently been claimed . That the observed failures of rational models are attributable to the cost of thinking and will thus be eliminated by proper incentives. Experimental findings provide little support for this view. Studies reported in the economic and psychological literature have shown that errors that are prevalent in responses to hypothetical questions persist even in the presence of significant monetary payoffs. In particular, elementary blunders of probabilistic reasoning (Grether, 1980; Tversky and Kahneman, 1983), major inconsistencies of choice (Grether and Plott, 1979; Slovic and Lichtentein, 1983), and violations of stochastic dominance in nontransparent problems … are hardly reduced by incentives" (Tversky and Kahneman, 1986, p. S274).

b) Even if we concede that anomalies can not be eliminated, it may be pointed out that individual deviations are *random* and, consequently, they will cancel each other at the aggregate level. However, as Thaler (1986) said, this argument fails because the biases are not only *common* but also *systematic*, a fact that allowed their classification in a variety of persistent "effects" whose occurrences can be predicted beforehand.

c) A frequent standard argument against the relevance of the "effects" is that agents can learn. Being aware that errors are costly they may learn and correct their choices in such a way that biases will be gradually eliminated. This was the kernel of Baumol's argument. Is it true that agents can learn from their mistakes and correct them? A moderate answer is "not always". Tversky (1974) and Tversky and Kahneman (1974) show that under uncertainty individuals make systematic mistakes in the formation of their probability judgments. The important point is that subjects keep behaving in a wrong way after being informed of the errors committed and instructed about how to cope with them. Einhorn and Hogarth (1978) argued in a similar way. It seems that learning isn't an easy task at all, and only takes place when relatively exceptional conditions are fulfilled.

> "The problem with many economic models of learning is that they seem to apply to a very static environment. In fact, such models seem to be directly applicable only to the situation in which Bill Murray finds himself in the movie *Groundhog Hog Day*. In the movie, Bill Murray is a TV weatherman sent to report on weather the groundhog sees his shadow on Feb. 2. Murray's character ends up reliving the same day over and over again. Although he is a slow learner, the opportunity to rerun the same day repeatedly, and to learn from the consequences of his actions each time, creates a controlled experiment in which he is able to learn many things eventually, from how to prevent accidents to how to play the piano. Alas, life is not like *Ground Hog Day*. In life, each day is different, and the most important of life's decisions, such as choosing a career or spouse, offer only a few chances for learning!" (Thaler, 1986, p.p. 135-136).[9]

d) Finally, it has been held (Friedman, 1953) that market mechanisms (arbitrage and competition) work in the right direction, eliminating the influence (and in the long run the very presence in the market) of those agents who behave in a sub-optimal way.

[9] Tversky and Kahneman (1986, pp. S274.S275) offer a more accurate classification of the many circumstances in which learning could not take place.

> "The argument proceeds something like this. Suppose there were some less-than-fully-rational agents. I like to call them 'quasi-rational', meaning trying hard but subject to systematic error. Once these quasi-rational started interacting with rational types, the rationals would quickly take all their money, after which the quasi's would either learn or would be rendered economically irrelevant" (Thaler, 2000, p. 136).
>
> The usual defense of the transitivity principle (money pump argument) and the probability axioms and Bayes's Theorem (Dutch Book Arguments) is based on this idea. However, "the claim that the market can be trusted to correct the effect of individual irrationalities cannot be made without supporting evidence, and the burden of specifying a plausible correcting mechanism should rest on those who make this claim" (Tversky and Kahneman, 1986, p. S275).

While the supporting evidence is still lacking, there are some indications that the opposite may be true. Russell and Thaler (1985) show that this is the case in the goods markets[10] and De Long et. al. (1999) make a similar point regarding the stock market. A tradition that goes back to Keynes explains the stability of the financial market by the unavoidable diversity of its players.

The many mechanisms just described, designed to explain away the anomalies failed to achieve their goal: the "effects" have been reproduced in different settings and their existence has been finally accepted. Theorists of decision making started then to elaborate alternative theories which allow the presence of the observed deviations. If there is a mismatch between theory and reality, and the facts cannot be modified the adjustment has to be achieved by changing the theory itself. However, EUT or SEUT were attractive mainly because of their *normative plausibility...* Could the alternative theories proposed for "explaining away" the phenomena retain this property?

5. The end of the Golden Dream

According to Kahneman and Tversky, EUT has two essential assumptions -*dominance* and *invariance*-, whose legitimacy from a normative standpoint seems at first sight unassailable. "Dominance is both simpler and more compelling that cancellation and transitivity, and serves as the cornerstone of the normative theory of choice", and invariance "captures the normative intuition that variations of form that do not affect the actual outcomes should not affect the choice" (Kahneman y Tversky, 1986, p. S253).

The problem arises because many choice situations show "framing effects", in which people's choices depend on how the options are described. This fact violates invariance, and

[10] "In this paper we start to explore the implications of irrationality for economics. … We then consider what happens if rational and less than fully rational agents … interact in the competitive markets. We show that knee-jerk reaction of some economists that competition will render irrationality irrelevant is apt only in very special cases, probably rarely observed in the real world" (Russell and Thaler, 1985, p. 1071). According to the authors the usual justification of maximizing models claiming that "markets guarantee that only rational behavior can survive" never applies, "except (perhaps) in some highly efficient financial markets" (id., p. 1080).

consequently dominance, because as Tversky (1999, p. 187) stated "any failure of description invariance can induce a violation of dominance".

It is important to grasp the nature of the difficulty pointed out. It is not just that agents as a matter of fact violate each of the assumptions. The trouble is that decision makers cannot, *in principle*, for *psychological* reasons, comply with the suppositions. According to Tversky and Kahneman (1986, pp. S256-S257) humans do not have mechanisms for securing invariance. As Kahneman (2003, p. 1459) explains:

> "The basic principle of framing is the passive acceptance of the formulation given. Because of this passivity, people fail to construct a canonical representation for all extensionally equivalent descriptions of a state of affairs (…). they do not spontaneously transform the representations of puzzles or decisions problems. Obviously, no one is able to recognize '137 x 24' and '3288' as 'the same' number without going through some elaborate computations. Invariance cannot be achieved by a finite mind".[11]

If they are right, the biases originated in framing effects are unavoidable given the cognitive and computational limitations of the human mind. The usual mainstream arguments according to which agents can be rational given satisfactory incentives and time enough for collecting information and learning are shattered. This virtuous mainstream process is a psychological impossibility. In our terminology, the assumption of invariance is *not feasible* for human beings. So, *no* adequate theory of *rational* decision can be satisfactory for *descriptive* purposes. One has to choose.

> "Because framing effects and the associated failures of invariance are ubiquitous, no adequate descriptive theory can ignore these phenomena. On the other hand, because invariance (or extensionality) is normatively indispensable, no adequate prescriptive theory should permit its violation. Consequently, the dream of constructing a theory that is acceptable both descriptively and normatively appears unrealizable" (Tversky and Kahneman, 1986, p. S272).

The usual vision that economists have of their own discipline has to be changed or, at least, qualified. It would not be permissible to say any more than:

> "the concept of rationality is used in economic analysis in three different ways: as a descriptive hypothesis about behaviour, as a normative concept, and as an aspiration, i.e., a way of organizing behaviour which is desirable and to which individuals and societies should be educated… [….] in fact the three points of view are in perpetual interaction" (Arrow, 1999, p. XIII).

6. How bad is the impossibility of the Golden Dream for economics?

Is it really fatal for an adequate *rational* decision theory the fact that it will be descriptively inaccurate? The dream of a fully satisfactory theory, reported at the beginning of this paper,

[11] From a very different perspective, Radner (2000) considers this point as a formidable setback for any theory of rational decision.

suggests that the answer should be YES. However, the ingenuity of Raiffa (1961) has brought about a counter-argument which is not only valuable in itself, but helps to throw light on the crucial distinction (frequently overlooked) between anomalies and paradoxes. In his view it is *not* desirable that a plausible normative theory of decision making happens to be also satisfactory from a descriptive standpoint, because if this goal were reached it would turn the theory completely irrelevant.[12] Raiffa's argument turns an evil into a virtue: the unfavorable empirical evidence regarding the pretension of truth of EUT or SEUT is converted into favorable evidence regarding their intended normative relevance. Besides, the Kahneman and Tversky argument about the supposed inability on the part of the agents for learning is dismissed. In fact, if people had the capacity for incorporating the rational decision rules of standard theory into his "intuitive system" and in so doing were able to decide rationally spontaneously and without effort, the theory would become descriptively right but its normative relevance would come to a halt. Against common sense Raiffa argues that anomalies are important because they testify about the relevance and necessity of the theory[13]. Crediting his ingenuity we will register this thesis with his name:

Raiffa's Law: The normative relevance of a theory of choice is directly related to its empirical failure (more precisely, to the amount and quality of the anomalies that overflow real choices)

In our view, however, Raiffa's law is not so strong as it seems. He misinterprets Ellsberg's point the same way Friedman and Savage misunderstand Baumol's objection. In both cases "anomaly" is erroneously identified with "paradox". Ellsberg is not only saying that it is possible to behave against Savage's axioms (which is trivially true), nor does he say that the individuals do commit in practice these violations. He claims that by performing "hypothetical experiments", choice situations may be identified in which it would be *reasonable* to choose in a manner that conflicts with the directions given by the standard theory. Ellsberg constructs a paradox, not just an anomaly. The important point is that Raiffa's law is irrelevant in the face of paradoxes. It may be true that the anomalies of the theory "reveal" its necessity as normative theory, but paradoxes put in question its credentials for doing the job.

The presence of a paradox involves the existence of an alternative intuitive notion of rationality which, at least in the particular situation considered, performs better than the standard one. If it is right that the presence of anomalies in decision making under risk or uncertainty shows that *some* normative theory is needed (not necessarily EUT or SEUT), it is also true that the presence of paradoxes involves the lack of credentials of the current theory for advising how to choose. To Raiffa's law we may oppose Ellsberg's law, in merit of having noticed the power of paradoxes when the target is the normative status of a theory of choice.

Ellsberg's law: The normative relevance of any theory of decision making is inversely related to the number and importance of the paradoxes that could be imagined.

To be aware of the differences between anomalies and paradoxes matters!

[12] "If most people behaved in a manner roughly consistent with Savage's theory then the theory would gain stature as a descriptive theory, but would lose a good deal of its normative importance. We do not have to teach people what comes naturally" (Raiffa, 1961, p. 690-691).
[13] If one gives credit to his idea, should it not be pushed a little further and claim that anomalies are welcome and their persistence convenient?

Bibliography

Arrow, K., Colombatto, E., Perlman, M., and Schmidt, M., (eds), (1999), The rational Foundations of Economic Behaviour – Proceedings of the IEA Conference held in Turin, Italy, Great Britain, Macmillan Press Ltd.

Baumol, William J., (1951), "The Neumann – Morgenstern Utility Index – An Ordinalist view", *Journal of Political Economy*, LIX, pp. 61 – 66.

Davidson, P. (1998) "Volatile financial markets and the speculator" (available at http://csf.colorado.edu/pkt/authors/Davidson.Paul/papers.html).

De Long, J. B., Schleifer, A., Summers, L., and Waldmann, R., 1990) "Noise Trader Risk in Financial Markets", *Journal of Political Economy*, 98, pp. 703 – 738.

Einhorn, H. and Hogarth, R., (1978), "Confidence in Judgment: Persistence in the illusion of validity", *Psychological Review* 85, Nº 5, pp 395 – 416.

Ellsberg, D., (1961), "Risk, Ambiguity and The Savage Axioms", *The Quarterly Journal of Economics*, Vol. 75, Nº 4, pp. 643 – 669.

Friedman, Milton and Savage, L. J., (1948), The Utility Analysis of Choices Involving Risk", *Journal of Political Economy*, LVI, pp. 279 – 304.

Friedman, M. and Savage, L., (1952) "The Expected Utility Hypothesis and the Measurability of utility", *The Journal of Political Economy*, Vol. LX, Nº 6.

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

Grether, D. and Plott, C., (1979) "Economic Theory and the Preference Reversal Phenomenon", *American Economic Review* 69: 623-638.

Hausman, D., (1992) *The Inexact and Separate Science of Economics*, Cambridge University press, Cambridge.

Kahneman, D, (2003a) "Maps of Bounded Rationality: Psychology for Behavioral Economics", *American Economic Review*, Vol. 93, Nº 5, pp. 1449-1475.

Kahneman, D., (2003b) "A psychological Perspective on Economics", *American Economic Review*, Vol. 93, Nº 2, pp. 162-168.

Kahneman, D. and Tversky, A, (1979) "Prospect Theory: An Analysis of Decision Under Risk", *Econometrica* 47, pp. 263-291.

Lichtenstein, S. and Slovic, P., (1971) "Reversal Preferences between bids and choices in gambling decision", *Journal of Experimental Psychology* 89, pp. 46-55.

Marschack, J., (1950) "Rational Behavior, Uncertain Prospects and Measurable Utility", *Econometrica,* Vol. 18, Nº 2, pp. 111-141.

Mongin, Ph., (1998) "Expected Utility Theory", in Davis, J., B., Hands, D., W., and Mâki, U., (eds) *The Handbook of Economic Methodology*, Cheltenham, Edward Elgar, pp. 171 - 178.

Radner, R., (2000) "Costly and Bounded Rationality in Individual and Team Decision-Making", *Industrial and Corporate Change*, v.9, no. 4, 623-658.

Savage, L., (1954) *The Foundations of Statistics*, New York: Wiley.

Schoemaker, P.J., (1982) "The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations", *JEL*, June.

Starmer, Chris, (2000) "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk", *Journal of Economic Literature*, Vol., XXXVIII, pp. 332-382.

Strotz, Robert H. (1953) "Recent Developments in Mathematical Economics and Econometrics: An Expository Session – Cardinal Utility", *American Economic Review*, Vol. 43, Nº 2, May 1953, pp. 384-397.

Thaler, Richard, (1986) "The Psychology and Economics Conference Handbook: Comments on Simon, on Einhorn and Hogarth, and on Tversky and Kahneman", *Journal of Business*, Vol 59, Nº 4, Part 2, pp. S279 – S284.

Thaler, Richard, (2000) "From Homo Economicus to Homo Sapiens", *Journal of Economic Perspectives*, Volume 14, Number 1, Winter 2000, pp. 133-141).

Thaler, R. and Russell, Th., (1985) "The Relevance of Quasi Rationality in Competitive Markets", The American Economic Review, Vol 75, Nº 5, pp. 1071 – 1082.

Tversky, A., (1974) "Assessing Uncertainty", *Journal of the Royal Statistical Society*, Vol. 36, Nº 2, pp. 148-159.

Tversky, A., (1999) "Rational Theory and Constructive Choice", in Arrow, K., Colombatto, E., Perlman, M., and Schmidt, M., (eds), *The Rational Foundations of Economic Behaviour*, Great Britain: Macmillan Press Ltd.

Tversky, A. and Kahneman, D (1974) "Judgment under Uncertainty: Heuristics and Biases", *Science*, New Series, Vol. 185, Nº 4157, pp. 1124 – 1131.

Tversky, A. and Kahneman, D. (1986) "Rational Choice and the Framing of Decisions", *Journal of Business*, Vol. 59, Nº 4, Part 2, pp. S251 – S278.

Raiffa, H., (1961), "Risk, Ambiguity and the Savage Axioms: Comment", *Quarterly Journal of Economics*, Vol. 75, Nº 4, pp. 690-694.

Von Neumann, J. and Morgenstern, O., (1944) *Theory of Games and Economic Behavior,* Princeton University Press.

Author contact: gustavoleomarques@hotmail.com and diego_mw@hotmail.com

SUGGESTED CITATION:
Gustavo Marqués and Diego Weisman, "Rethinking Rationality Theory's epistemological status. Normative vs Positive approaches", *real-world economics review*, issue no. 77, 10 December 2016, pp. 50-63,
http://www.paecon.net/PAEReview/issue77/Marques-Weisman77.pdf

real-world economics review, issue no. 77
subscribe for free

Time and the analysis of economic decision making[1]

Donald W. Katzner [University of Massachusetts/Amherst, USA]

Abstract
Economics is concerned, to a considerable extent, with explaining real economic behavior. And real economic behavior is generated by the making of real economic decisions. Both real economic behavior and real economic decision making occur in real time. How real time is represented, indeed if it appears explicitly at all, reflects significantly on the realism and relevance of the explanation produced by an analysis. This paper explores the role of real time in the analysis of economic decision making through its impact on the knowledge and ignorance that is present when decisions are made.

Introduction

Economics is concerned, to a considerable extent, with explaining real economic behavior. And real economic behavior is generated by the making of real economic decisions. Both real economic behavior and real economic decision making occur in real, or what is often referred to as historic time. Real time, then, is implicit if not explicit in all economic analysis. How real time is represented, indeed if it appears explicitly at all, reflects significantly on the realism and relevance of the explanation produced by an analysis.

The concept of time, in and of itself and apart from the meaning and implications of time in economic analysis, economic policy proposals, and economic conduct, has puzzled philosophers as to its ontological status and the epistemological challenges implicit in it. Sir Thomas Browne, while admitting to the inability of solving the riddle of time, nevertheless observed in his *Hydriotaphia*: "Time which antiquates antiquities . . . hath an art to make dust of all things".[2] Time passes. History occurs. And its deposits too frequently falsify our expectations. Browne's failure to solve the riddle of time, though he was apparently aware of its often unforgiving imperatives, echoes both subsequent and prior philosophic investigations. Eric Rosenfield has observed to similar effect:

> "In 1917, Albert Einstein completed work on the *General Theory of Relativity*, one of the rules of which states that time is fundamentally bound to matter and gravity. Oddly, this concept was presaged almost 1,300 years before when Augustine (in Book 11 of his *Confessions*) put forth the idea that when God created the Heaven and the earth he created time as well." [3]

Human beings experience real time as a series of durations – not as a sequence of isolated points lined up along a continuum.[4] Durations do not have a well-defined length. They do not follow each other as separable and discretely identifiable intervals with a precise beginning and a precise end; each overlaps is predecessor and successor. Moreover durations are frequently perceived with respect to the events they contain. That is, time is often felt in a vague sort of way as events come into view, evolve, and pass into history. With this in mind, it

[1] The author would like to thank Douglas Vickers for his considerable help.
[2] Browne (2012, Ch. V).
[3] Rosenfield (2015, p. 7).
[4] Georgescu-Roegen, (1971, p. 70).

suffices here to take account of the ontological status of time as an entity in itself that moves and rolls as a stream and alters human consciousness between yesterday, today, and tomorrow. From such a perspective, time presents the challenge of knowing, with significance for human action, what it contained yesterday, what it means for action and behavior today, and what it might encompass tomorrow. And this, in turn, implies that the ontological status of time throws up profoundly significant, perhaps insoluble epistemological questions. Its significance for economic argument and conduct protrudes on several levels that have to do with the interdependent realizations of knowledge and ignorance. Three such issues immediately raise their demands.

First, it was Mill (1874, Sect 3, p. 237) who wrote that the cause of any event is the entire "… set of antecedents which determine it, and but for which it would not have happened." In the present context, that statement points to the following question: to what extent is it possible to know enough of what has occurred in the past and the causes of those occurrences in order to determine adequately the possibilities of human action in the present? That in itself, as will become clear as present argument proceeds, is not simply a statistical problem. For involved in it are matters of human epistemic potential, including in that the nature of presuppositions that influence one's look into the past, and what it is that determined those presuppositions. The limits of epistemic finitude exert their sway. We face the problem of how do we know, what are the origins and processes of knowledge, and what are the validity criteria that vindicate our grasp of what we conclude occurred in the past. That nexus of inquiry takes up also one's realization of the inter-determining forces and relationships that cause the past to be what we suppose it to have been. For example, is one to take a severely deterministic view of the past, is it to be understood, as some contemporary postmodern arguments suppose, to be the depository of chance events and outcomes, or is the past capable of bequeathing distinctly cognizable lines of explanation?

Knowledge of the past emerges, of course, from experience over time – that is, observing, reading, and interacting in other ways with individuals, groups, and the environment. Combining those activities with thought processes about what, with varying degrees of certitude, is already known produces new interpretations and possible bases for action in the future. All of this is stored in memory and added to knowledge. But there are pitfalls: We are unable to observe and read everything; nor can we know everything from the past. Thus there are necessarily big gaps in our knowledge. Our interpretations and understandings of observations, readings, and interactions may not be accurate. And as time passes, memories dim. We forget some things, we often change our understandings or meanings of others, and we add some that were never present. Time, therefore, alters knowledge by making it, at the same time, both more and less reflective of the world in which we live and of our assumed history of that world.

Second, what, as result, is it possible to understand as accessible grounds of responsible decisions for action in the present? That is only partly dependent on what might have been resolved by the discussion of issues raised in the preceding. For now there comes into play certain highly significant matters regarding the manner in which a marriage might be effected between two considerations raised by answers to the previous questions; first, what is now to be understood to be the history that precedes human action; and second, what can be discerned as the predilections, preferences, and predispositions that interact to determine what is seen as contemplated and desired results of present decisions. The latter, of course, is the goal-oriented grounds of action or criteria on the basis on which the decision choice is made. Indeed, what shall be suggested as possible decision criteria will take heavily into

account a number of such epistemic considerations. In any case, it is the union of these two considerations that produces the foundation for the decision.

Apart from knowledge in time that goes into the making of decisions, the cultural and social backgrounds of the decision maker are also highly significant.[5] That is because decisions emerge from more or less ordered and articulated thought processes. Thought processes, in turn, are mental acts that rely heavily on the symbols and their interpretations that individuals import across time as part of their intellectual maturation and subsequent development. Motivations are included in those imports. The source of the imports is the cultural and social environment in which the individual lives, acts, and grows. Thought processes that result in decisions involve, in light of motivations, the manipulation of the imported symbols in culturally and socially determined ways. It follows that since cultural and social values and imperatives evolve over time, the thought processes of the individual move in (possibly lagged) parallel fashion. The passage of time, then, may lead to variation in decision-making thought processes and the consequent alteration in decision outcomes from what might otherwise have been.

Third, in its reluctance to confide to us a full understanding of what occurred in its past, and in our incomplete comprehension of the meaning of the present, unarguable as that is, time hides from us completely what it will disclose in the future. For the epistemic reality we face is not simply that the future is unknown. It is unknowable. It would be an epistemic trap that leads to a blind alley to imagine that it is possible to know a part, even if in humility we say only a part, of what the future will disclose. The realities with which the inexorable passing of time has to be cognized are that as we look from our present posture to the future we are ignorant of what it will contain.[6] It might be imagined that on the basis of what we have constructed as the events of the past, we can entertain certain expectations or hunches of what the future may bring. Part of what will be observed below as decision criteria and ways of corralling the unknown and unknowable future, have been based on some such notions. But it must be acknowledged that all we have is our 'construction' of the past, never, as referred to above, a true, in the sense of accuracy and completeness, knowledge of the past. But therein lies a significant part of the problem that confronts the economic analyst or forecaster or decision maker. And for that reason the epistemological questions raised here will influence the decision processes and criteria that will be subsequently examined.

The argument that follows will look, in intentionally brief and incomplete terms, at the ways in which the already-addressed issues emerging from the acknowledgment of real time are relevant to the analysis of decision-making in economics. It is concerned, then, with first, the inability to know the past completely and comprehensively; second, the possible relation between that imperfect knowledge and the particular preferences, predilections, and presuppositions that influence the criteria of the decision maker's action; and third, the ignorance that confronts the decision maker as he looks into the unknowable future. The future, it needs to be grasped, is not there to be observed in advance. Decisions made in the present create the future that emerges.

Economic analysis, in its historically long development has, of course, been conscious that the realities of time and its passing need to be taken into account. As economics as an

[5] See Katzner (2008, pp. 5, 6 and Essay 3).
[6] As will be argued below, we cannot even know the probabilities of future events. The economic world is decidedly non-Bayesian.

intellectual discipline matured, something of an antagonism existed between opposing views of the manner in which this was to be accomplished.[7] Adam Smith, in his insightful inquiry into the nature and causes of the wealth of nations, looked essentially at the ways in which, in actual time, the relations inherent in market activity did, in fact, work out to the mutual advantage of participants. But real time as such was not explicit as a determining factor or variable in his analysis. In the classical economics that followed Smith, and notably at the hands of David Ricardo, the problem of real time was essentially solved by assuming it away. Ricardo's approach was based on the supposition that if left to itself the market system would automatically lead to a full employment of economic resources and, as a result, to maximum attainable economic welfare. The classical economists, in eliding the problem of, and the possible disturbances resulting from, the passing of real time, imagined that the economic system was shot through with automatic harmonies.

But so far as the explicit recognition of time was concerned, all that changed as the nineteenth century progressed and gave rise to what became referred to as neoclassical economics which, in turn, projected its analytical content into the twentieth and twenty-first centuries. Now the question of time was, in a special sense, recognized and taken into account. On more levels of analytical sophistication than can be addressed in the present space, distinctions were contemplated between the short run and the long run in economic affairs and outcomes, as they depended on individuals' decisions. But the time-scale differences did not address what we are now referring to as real historic time. Analytical focus remained on what is often referred to as logical time.[8] Decisions now took place in that time, and the imagined length of time over which resulting outcomes were contemplated was accorded a very important place in the analysis. Indeed, on such a logical level, progress was made in the dynamic view of things and time-paths over which the economy may develop, with convergent or explosive results and corresponding equilibrium or disequilibrium postures extensively examined. But nevertheless, time was in all that conceived of as a logical variable. Underlying that development in analysis were assumptions imported from the preceding classical economics that in the long run the market system would, in the general case, automatically lead to full employment of economic resources and again to maximum attainable economic welfare. At the base of the analysis were assumptions that the market system was characterized by perfect competition between small firms in various markets, producing identical products, under conditions of perfect knowledge, automatic and perfectly rapid market adjustment mechanisms, and assumptions of freedom of entry to, and exit from, markets. All that was subject to modification as time moved on, of course, and consideration was given to forms of imperfect competition such as duopoly and monopoly, and assumptions of decision makers' alternative actions were introduced.

The point that argues for attention, however, is that throughout that analytical evolution, important significance attaches to the fact that time was taken as simply a logical variable in the analysis. When, in such ways, time is simply a logical variable, the differences between yesterday, today, and tomorrow can be introduced into the analysis without any recognition that the real passing of time can, and does in fact, cause changes in the various variables inherent in the analysis of the scheme of things. At that juncture, if the question of real time is

[7] See Vickers (1994).
[8] When time is viewed as logical, pairwise comparisons of events are made in such a way that any one event, regardless of whether it occurred in the past, occurs in the present, or is thought to possibly occur in the future, is only said to take place before, simultaneously with, or after another. All that is important is the sequencing of events. The specific and distinguishing qualities of being in the past, present, or future are irrelevant and ignored.

to be taken adequately into account, there is a fly in the ointment. The awareness of it is at this point critically important when we are dealing with analytical economics and the relevance of actual decisions within it, and not merely with descriptive economics. For in systems of logical time analysis there do not really exist any choices that a decision maker in real time would face in contemplating the movement of the economic system from one time point to another, from the present to the future. A subtlety exists to spoil the argument. For when real time is taken into account, it is difficult to agree, as has generally been stated, that economics is concerned with nothing more than choices between achievable ends by the allocation of present (scarce) resources. Rather, if the general assumptions of neoclassical economics as have previously been indicated are present in the analysis, the outcome is automatically determined by the assumption content of the thought-system that is in place, and no real choice or actual decisions exist. The analytical structure of the system determines its own outcome. On the contrary, genuine choice in real time exists only when real uncertainty is present. That is so because in that case judgments have to be made, real decisions and real decision responsibilities exist, and the entire compass of unknowable possible outcomes have to be contemplated.

Real time, then, ignorance, and responsible judgment provide a context for the analysis of economic decision making that propels thought in significantly different directions from what economics has, in the past, substantially supposed. What needs to be considered now, therefore, is a brief recognition of the ways in which economic analysis has, with possible degrees of reality and pretense, taken real time into account.

To begin that argument it is necessary to recognize an essential difference between two very different methods of proceeding. First, by the use of certain analytical tools and assumptions, most usually the introduction of the probability calculus, the future with its ignorance and residual uncertainty can be and is, in effect, assumed away. Future-dated variables, subjected to estimation so as to be describable by the first and second (and possibly higher) moments of assumed probability distributions, may be reduced to present values by appropriate discounting procedures. The assumption of complete contracts – that the parties involved are able to specify commitments in every possible future state – follows a similar path. Such ways of incorporating the significance of real time into the analysis have to be seen as essentially attempts to transform ignorance into knowledge, opening serious argument as to the epistemic grounds on which, with the future unknowable in real time, the underlying probabilities inherent in the analysis are capable of specification. More will be said of this below.

It is, of course, possible that history might give the appearance of ‘repeating’ itself. That is, given our limited knowledge of past and present, it may seem that the occurrences of one day were identical to those of the previous day. But such repetition can only be observed after both days have passed in time. Moreover, in the context of decision making under the supposition of repetition in which all conditions of the decision problem were assumed to remain fixed, no account could be taken of the fact that real time passes between the making of the decision and the realization of the outcome of the decision, and hence that the world is in fact changing as the decision outcome is generated. The actual outcome of the decision, then, could be quite different from what might have been anticipated. Even with historic repetition, there is still no epistemic basis for the assumption that the same conditions will be repeated again tomorrow. History, as has been said, is created from moment to moment. Only time can reveal the mysteries of what is to come.

In the second approach to the making of economic decisions, the force and significance of real time is present throughout. Specifically, time enters the analysis by "… taking note of the manner in which the actual flow of it, and the unknowable expanse of it spread out ahead of us, impinge on real-world choices …. [T]ime is significant because our imaginative perceptions of the possibilities inherent in it determine what we do in our choice-decision moments, and because the passing of time qualifies our stance at the decision points we confront".[9] An explanatory model of decision making which to a considerable extent accounts for time in this way was proposed by Shackle (1969) and modified by Vickers (1987, Ch.12). The remainder of the present paper is concerned with the appearance and role of real historic time in that model. The discussion that follows is heavily dependent on the present author's expanded discussion in his *Time, Ignorance, and Uncertainty in Economic Models* (Katzner, 1998) of the analysis of decision making under the Shackle-Vickers assumptions of real time and ignorance of the future.[10]

It is natural to begin with the objects among which a decision or choice is to be made. In the concrete example presented below, those objects are alternative investment capital outlays. More generally, the objects of choice are known to the decision maker with the accuracy and completeness that the acquisition of knowledge over time as previously described permits. That is not to say that, at some future date, it might not become clear that other choice objects, unknown to the decision maker at the decision point, might actually have been available. However, the possibility of making adjustments to the decision as the future unfolds will not be considered here.

Upon making a decision, the selected object of choice or the elements of the choice decision interact with future states of the world that they meet as time passes to produce the outcome of the decision that was made. Because those states arise after the decision is made, perhaps both while and after it is being carried out, they are unknowable at the decision point. But the states of the world that greet the decision have a considerable impact on its outcome, and whether the decision made may, in retrospect, be regarded as a success or failure. The uncertainty of future states created by the veil of time should somehow be recognized in the decision-making process.

Expanding on what has earlier been said, many economists account for this uncertainty in terms of probability, that is, by assigning probabilities to states of the world. Probability arises in two forms: Aleatory probability is associated with the outcome of chance mechanisms and the relative frequencies they produce upon repeated trials in unchanging environments; epistemological probability is concerned with measures of degrees of belief, as warranted by evidence or reasonably informed judgment, that outcomes will obtain.[11] But both forms of probability leave something to be desired in their encounters with the difficulties of real time in the context of economic decision making. On the one hand, aleatory probability fails on two grounds. First, to calculate frequencies of states of the world requires, in part, knowledge of all possible state outcomes that can be produced by whatever chance mechanism is thought to generate them. But since the future is unknowable, it is not possible to be cognizant of all future states even if one allowed for the possibility of accurate knowledge of all states that have come before. In addition, each moment as time passes is unique in the sense that it has a unique history of actions and happenings, and unique collections of individuals and

[9] Vickers (1994, pp. 194-195).
[10] See Katzner (1998, Sect. 4.3).
[11] Hacking (1975, Chs. 1, 2).

institutions. The latter have unique preferences, attitudes, values, etc. that have evolved in the past and will modify in the future. Such uniqueness precludes the repetitive trials in unchanging environments that would generate frequencies of appearances of states of the world. On the other hand, calculating the epistemological probability of future states of the world also leaves something to be desired because the knowledge necessary as evidence relating to futures states or to make the reasonably informed judgments about them cannot, because the future is unknowable, be present.

An alternative way of assessing the uncertainty of futures states of the world is in terms of degrees of surprise that might be held regarding them. Surprise is based on imperfect knowledge of the past and on current psychological structures as they relate to such elements as attitudes and predilections. It does not require a leap across time into speculations regarding the actions, happenings, and nature of individuals and institutions of the unknowable future. And it is applied only to what are considered to be possible future states, recognizing that other states, currently unknown, may well appear as the future unfolds. The specific notion as originally proposed by Shackle (1969, pp. 68-70) was called potential surprise. It is defined as follows: The potential surprise of a possible future state of the world is the surprise the individual imagines now that he would experience in the future if that state were actually to come to pass. This notion captures to a sufficient extent the uncertainty in the economic decision problem and requires no unknowable knowledge of future states of the world.

A given choice object will be connected in the decision maker's mind to a range of possible outcomes, and to each such outcome he will attribute a potential surprise value. That, then, provides a set of pairs consisting of possible outcomes and the degrees of potential surprise with which they are contemplated. Outcomes could be expressed, perhaps, in terms of utility or profit. In either case, some outcomes, say those identified with higher utility or profit, would be seen as more favorable than others. Now for the given choice object, looking on the more favorable side, the decision maker's attention is pulled towards one particular combination of a more favorable outcome and a potential surprise value associated with it. That is, given the decision maker's psychological make-up at this moment of time, he is in some way drawn or attracted to that combination. Note that this combination of a more favorable outcome and a potential surprise value is one member of the previously described set of possible outcomes and the potential surprise values to which they relate. The analytical details concerning the manner in which the decision maker comes to focus on this particular combination will be outlined in the illustration below.

Similarly, on the less favorable side, a different combination stands out in the decision maker's thoughts or attracts his attention. These two combinations may be used to fully characterize his view of the choice object in question. Thus the entire decision problem is reduced to the selection from a collection of choice objects, each represented by two combinations that have grabbed the decision-maker's attention (consisting of a more favorable outcome and a less favorable outcome with their associated potential surprise values), according to a specific criterion in the decision maker's mind. It is assumed that that criterion is to select a choice object whose representation by the two combinations is the highest on what may be referred to as a decision index.[12]

[12] See Vickers (1987, pp. 222-235).

The explanatory model of decision making just outlined (i.e., that of Shackle and Vickers) has a more complex structure than those based on the frequently-employed assumption that future possible outcomes can be specified probabilistically, often making use in such a process of the statistical moments of relevant probability distributions. But its focus on potential surprise, attractiveness, and the decision criterion is intended to follow the possible thought process of the decision maker making his decision. In that regard, the known objects of choice are first captured in terms of the surprise they may call forth and their attractiveness to the decision maker. To reach a decision, the decision criterion is then applied to the manifestation of these characteristics in each choice object. The process does not rely on a construct, namely probability, that requires knowledge never available to the decision maker.

It will be useful to conclude with a concrete example of a decision-making situation as that might be conceived of in terms of the potential surprise approach. First, let it be imagined that an entrepreneur faces the possible wisdom of investing distinct designated sums of capital in an expansion of an industrial plant. That is, there are different possible expansion projects, each requiring a unique investment or quantity of funds. In terms of the proposed decision procedure, for each project or possible investment the entrepreneur would contemplate a range of possible profit outcomes (or present capitalized values of those outcomes), and to each such outcome he would assign a potential surprise magnitude. That is, to recall, he would identify the degrees of surprise that he thinks now he would realize at a designated future date if particular possible results were, in fact, to occur. In connection with each imagined investment project, then, he would have recognized in his thinking a set of pairs of (i) possible outcomes and (ii) the potential surprise magnitudes associated with them. For each such project, the elements of that set can be understood to describe a potential surprise function defined over the domain of possible profit outcomes from the contemplated investment. It should be borne in mind that the domain of such a function would include the range of negative or loss (less favorable), as well as positive or gain (more favorable) outcomes. That is to say, possible financial losses as well as possible financial gains will occur to the decision maker as conceivable.

Second, and independently of any of the investment projects, the decision maker will, in his estimation of things, conceivably be prepared to compare combinations consisting of the various possible profit outcomes and their corresponding potential surprise values according to their attractiveness to him. The set of relations between such contemplated profit and surprise combinations, then, can be interpreted to mean that the decision maker holds in mind what might be called an 'attractiveness function,' defined over the Cartesian product of combined negative and positive ranges of possible outcomes and the range of potential surprise values. Such a function indicates the combinations of profit outcomes and potential surprise values which he would consider equally attractive. What is in view at that stage is akin to the well-known utility function in, for example, the neoclassical theory of consumer commodity choice. In the same way as with the familiar utility function, equally attractive pairs (of profit outcomes and potential surprise values) would be taken to describe iso-attractiveness contours in the profit-potential surprise space. Clearly, the realities of entrepreneurial investment require it to be realized by the decision maker that such iso-attractiveness contours will be described in the negative or loss quadrant, as well as in the positive or gain quadrant.[13] The contours in the two quadrants are independent of each other. When the iso-attractiveness contours and the potential surprise function from a specific

[13] The loss quadrant is defined as the Cartesian product of the range of loss outcomes and the range of potential surprise values. The gain quadrant is defined in similar fashion.

investment project are brought together, it will emerge that the respective functions osculate at a unique pair in the positive or gain quadrant, and similarly at a unique pair in the negative or loss quadrant. At those points of osculation (given appropriate assumptions of functional forms) the potential surprise function will touch in each case the highest achievable contour of the attractiveness function. It is to these unique pairs the decision-maker's attention is drawn. Then the desirability of the capital investment outlay that is in view can be located on a decision index constructed by taking account of both attention-grabbing positive and negative pairs. A point on the decision index can be observed at a magnitude determined by assigning values to the combination of those gain and loss pairs, conceivably assigning a negative value to potential unfavorable outcomes, and a positive value to potential favorable outcomes.

What has just been described as a procedure to assess the desirability of the investment project referred to at the beginning may be repeated with respect to every other possible investment project facing the decision maker. In that way, when all possible alternative investment project outlays that are available and competing for the entrepreneur's capital funds are brought to comparison on the same decision index, the entrepreneur can choose between various projects and construct his economic organization according to that which registers the highest or most prominent.

It should be clear that all of the structural components of this decision process are relevant only at the decision moment. As time moves on and the future comes into view, the emerging state of the world and the outcome of the decision take their place in history. And at the next decision point, the presumption is that the structural components are all different and work their way through the decision process to a different decision outcome.

References

Browne, Sir T. (2012) "Hydriotaphia," *Religio Medici and Urne-Burial*, S. Greenblatt and R. Targoff, eds. New York: New York Review of Books.

Georgescu-Roegen, N. (1971) *The Entropy Law and the Economic Process*. Cambridge: Harvard University Press.

Hacking, I. (1975) *The Emergence of Probability*. Cambridge: Cambridge University Press.

Katzner, D.W. (1998) *Time, Ignorance, and Uncertainty in Economic Models*. Ann Arbor: University of Michigan Press.

Katzner, D.W. (2008) *Culture and Economic Behavior*: *Economics in the US and Japan*. New York: Routledge.

Mill, J.S. (1874) *A System of Logic*: *Ratiocinative and Inductive*, 8th ed. New York: Harper.

Rosenfield, E. "An Analysis of the Concept of Time in the *Confessions*, Book 11 of Augustine of Hippo," p. 7. Accessed at http://the-wanderling.com/augustine_time.html on 9/2/2015.

Shackle, G.L.S. (1969) *Decision, Order, and Time in Human Affairs*, 2nd ed. Cambridge: Cambridge University Press.

Vickers, D. (1987) *Money Capital in the Theory of the Firm*. Cambridge: Cambridge University Press.

Vickers, D. (1994) *Economics and the Antagonism of Time*. Ann Arbor: University of Michigan Press.

Author contact: dkatzner@econs.umass.edu

SUGGESTED CITATION:
Donald W. Katzner, "Time and the analysis of economic decision making", *real-world economics review*, issue no. 77, 10 December 2016, pp. 64-72,
http://www.paecon.net/PAEReview/issue77/Katzner77.pdf

real-world economics review, issue no. 77
subscribe for free

The mathematical equivalence of Marshallian analysis and "general equilibrium" theory

Philip George [India]

Abstract
In this paper we explore general equilibrium (GE) theory. We demonstrate that (a) Marshallian demand analysis is no less general in its implications, though this is masked by the way static analysis is framed and (b) that an unstated assumption of GE theory is that aggregate demand is constant. Together, these two results imply that, shorn of the complicated math, GE theory is equivalent to Marshallian demand analysis. It also explains why the two arrive at identical results on subjects like involuntary unemployment. This in turn, has implications for how we view the work of Keynes as well as current issues in economics, such as the legacy of formalism, and the relevance of new classical economics, dynamic stochastic general equilibrium (DSGE) and real business cycle theory.

Keywords: general equilibrium theory, Kenneth Arrow, macroeconomics, Marshallian economics

JEL Classifications: E12, E13

Introduction

A form of Alfred Marshall's work is still present in general textbooks and historians of economic thought continue to debate the significance and achievements of Marshall. On this basis it is perhaps possible to distinguish between "Marshallian" as an interpretation and use of his legacy and the context, intentions, limits etc. of his *Principles*. For a long time now it has been customary for economists to think of Marshall's work as a relic from a mathematically-challenged era. This is in so far as static analysis for partial equilibrium is superseded by general equilibrium (GE) theory. It is now also widely recognized that GE theory is problematic in various ways. Modern mainstream economics cannot be reduced to it. However, it remains influential both as a point of reference for theory and as a standard for formalism as an important mode of expression within economics. Moreover, GE has been highly influential during the period in which Keynes' work and the concept of involuntary unemployment have been distorted or suppressed. With these issues in mind, in the following short paper I make three important points:

1. Marshallian demand analysis as partial equilibrium necessarily involves effects on all markets and so is equivalent to GE;
2. GE theory does not disprove the existence of involuntary unemployment, rather it assumes it out of existence. Notably, aggregate demand is assumed to be constant.
3. However, the most reasonable explanation of unemployment within GE is still always likely involving involuntary unemployment, due to the cumulative impact of even small disequilibria on the special (if unacknowledged) status of the labour market as the largest (and in a generalised sense pervasive) market.

The underlying point made is that GE is not a significant improvement on partial equilibrium. It is rather part of the way in which Keynesian approaches have been disadvantaged or

distorted. Keynes' work remains relevant. Keynes' own words in the *General Theory* on the battle between Ricardian and Malthusian ideas seem relevant here:

> "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 other point of view completely disappeared; it ceased to be discussed" (Keynes, 1936).

The argument is made drawing heavily on Arrow's work on GE, since Arrow is careful to state the key assumptions. Since the paper is brief, and the aspects of Arrow's work drawn on relate closely to mathematical expression, extensive sections of his work are reproduced here.

A. The generality of general equilibrium theory

One of the principal claims of GE theory is that it takes all markets into simultaneous consideration and thus more realistically reflects the interconnectedness of all markets in the economy than Marshallian economics, which deals only with one isolated market at a time and is thus a case of partial equilibrium.

However, Marshallian demand analysis too deals with the economy as a whole and is thus no less general than GE theory. This is not typically acknowledged within Marshallian analysis, though, as I will show, the assumptions made in the construction of each market have consequences for all others. Clearly, GE differentiates itself on the basis of the actual intent of the theory, but this does not mean the two lack equivalence.

Kenneth Arrow's Nobel Memorial Lecture (Arrow, 1972) sets out the GE argument:

> "The consumer starts with the possession of some quantities of economically valuable goods, such as labor of particular types, land, or other possessions. Let us imagine there are n commodities altogether, and let $\bar{x}_{hi}$ be the amount of commodity i owned initially by individual h (this may well be zero for most commodities). If p_i is the price of the i^{th} commodity, then his total income available for expenditure is
>
> $$\sum_{i=1}^{n} p_i \bar{x}_{hi} \qquad (1)$$
>
> "Hence, he can choose for consumption any bundle of goods, $x_{h1}, \ldots, x_{hn}$, which cost no more than his income,
>
> $$\sum_{i=1}^{n} p_i x_{hi} \leq \sum_{i=1}^{n} p_i \bar{x}_{hi} \qquad (2)$$
>
> "Within this budget set of possible consumption bundles, the individual is presumed to choose his most preferred bundle... The most preferred bundle then is a function,
>
> $$x_{hi}(p_1, \ldots, p_n) \qquad (3)$$

> of all prices. Notice that, from this viewpoint, all prices clearly enter into the determination of the demand for any one commodity. For one thing, the rise in any one price clearly diminishes the residual income available for all other commodities. More specifically, however, the demands for some commodities are closely interrelated with others; thus, the demand for gasoline is perhaps more influenced by the use of automobiles and therefore by their price than it is by its own price. The interrelation of all demands is clearly displayed here" (Arrow, 1972).

The idea that general equilibrium theory is superior to alternative methods is clearly stated in the subsequent literature. For example:

> "The Walrasian [general equilibrium] theory has the capacity to explain the influence of taste, technology, and the distribution of wealth and resources on the determination of value. Nothing that came before the Walrasian theory had this capacity. Neither partial equilibrium theory nor theories that depend on technology and resources alone provide as strong an explanation of value. Although, for certain markets, it is possible to explain how price responds to small parameter changes with partial equilibrium reasoning, few economists would contend that this method is adequate when economies are disturbed in a major way" (Duffie and Sonnenschein, 1989).

Or again, a little later:

> "The essence of general equilibrium does not preclude aggregation; what is essential is an emphasis on inter-market relations and the requirement that variables are not held fixed in an ad hoc manner" (Duffie and Sonnenschein, 1989).

We turn next to Marshallian demand curves to see if this claim to significant difference is true.

Figure 1 below shows a linear demand curve for the fish markets.

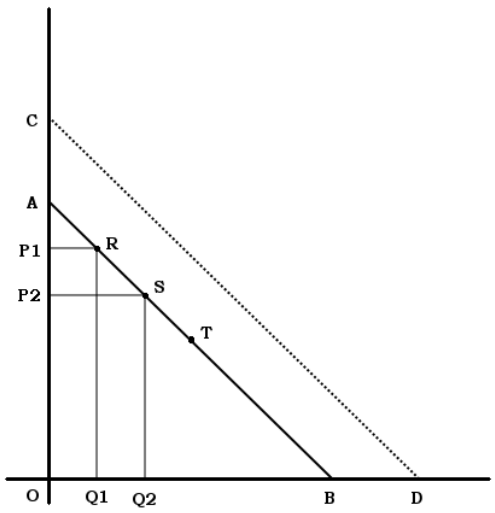


Fig 1. Linear demand curves for fish

The original demand curve is AB. The demand curve is drawn assuming that people's incomes are constant as are their tastes. If people's incomes increase then at every price they can buy more fish and the demand curve moves to CD. Similarly, if people develop an increased taste for fish the demand curve moves up.

T is the midpoint of the demand curve. The segment AT is the elastic zone. The segment TB is the inelastic zone.

Assume that the initial equilibrium is at point R (in the elastic zone) where the price is P1 and the quantity sold is Q1. Assume also that at this point individuals spend all their income and do not save anything. Next suppose that because of a movement of the supply curve the price falls to P2, also in the elastic zone, as a result of which the equilibrium moves to S. The quantity of fish bought increases to Q2. We can also see from the graph that the money spent on fish rises; the initial amount spent is the area of OP1RQ1 and the final amount spent is the area of OP2SQ2.

But here we run into a problem. We had assumed that individuals spent all their income at the first equilibrium point R and now we find them spending a larger amount on fish at S. This can happen only if they spend a smaller amount on some other good or goods so as to maintain their spending constant.

To summarize, along the elastic portion of a linear demand curve, when the price of fish falls not only does the quantity of fish bought increase but the money spent on fish also increases. So the money spent in other markets has to fall so as to maintain our initial assumption of a constant income. In general, the money spent at any point on the demand curve is different from that spent at any other point. To compensate for this difference the money spent in other markets, and therefore the demand and price in those markets, have to change.

This feature of linear demand curves also applies to demand curves of other shapes, with a solitary exception: the rectangular hyperbola PQ = constant. An analysis involving the rectangular hyperbola is too complex to be gone into here.

What is true of the fish market is also true of *every other* Marshallian market. When the price and demand for any good changes it affects demands and prices in other markets.

So it is clearly not the case that the Marshallian demand curve is drawn on the assumption that demands and prices in all other markets are constant. The Marshallian analysis is, therefore, no less general than General Equilibrium analysis. The charge is that in the Marshallian analysis, the individual's demand for the i[th] commodity is $x_{hi}(p_i)$ whereas in reality it is $x_{hi}(p_1, \ldots, p_n)$ which is the demand for good i in GE theory. The two are in this sense, other claims apart, equivalent.

B. General equilibrium theory and involuntary unemployment

In his Nobel lecture Arrow describes the various stages by which GE theory arrives at an equilibrium that is also Pareto efficient. But after the proof is done he mentions a caveat:

> "There is one loose end that should now be picked up. It has been assumed that the demand functions of the individual are continuous. But one of the surprising discoveries that [Gerard] Debreu and I made in the course of our study was that even under all the usual strong assumptions about the behavior of individuals, this cannot be true everywhere in the price simplex except under very artificial conditions. The trouble is that the individual's income also depends upon prices, and if the prices of those commodities which the individual owns originally fall to zero, his income falls to zero. When some prices and income are zero, however, the demand for the now-free goods may jump discontinuously. To illustrate, suppose an individual owned initially only one good, say, labor. So long as the price of that good was positive, he might retain some for his own use, but in any case could never consume more than he had initially. But when the price fell to zero, he could demand the same labor from others and in any amount he chooses. The existence of competitive equilibrium then does depend on assumptions which insure that for each individual there is at least one commodity he owns initially which is bound to have positive value" (Arrow, 1972).

Duffie and Sonnenschein's paper provides further elaboration. They quote Arrow:

> "Debreu and I sent our manuscripts to each other and so discovered our common purpose. We also detected the same flaw in each other's work; we had ignored the possibility of discontinuity when prices vary in such a way that some consumers' incomes approach zero. We then collaborated, mostly by correspondence, until we had come to some resolution of this problem." D&S go on to explain: "This resolution was to require, in theorem 1 of their paper, that the initial endowment of each household be interior to its consumption set. (Arrow had faced a difficulty much related to the demand discontinuity problem in his earlier work on the second welfare theorem)" (Duffie and Sonnenschein, 1989).

But if the existence of a general equilibrium requires ruling out a situation when a consumer's initial endowment has a market price of zero it *also* requires ruling out a situation when the consumer possesses only a single commodity and is unable to sell it because this would practically mean that he had no initial endowment. When the commodity in question is labour the situation is, of course, what we call involuntary unemployment. That is to say, labour has a market price but not all consumers who possess it can find a buyer.

So, contrary to what at least some GE theorists would have us believe, GE theory *does not disprove the existence of involuntary unemployment*. Rather, for GE to exist, *involuntary unemployment must first be assumed out of existence*. General equilibrium can exist only if its assumptions guarantee that the labour market has no disequilibrium. And yet GE is also an equilibrium seeking theory. Marshallian demand analysis logically arrives at the conclusion that involuntary unemployment is impossible, but GE theory needs to assume it in the first place to ensure general equilibrium. The equivalence between the two approaches thus involves some difference, but the difference does not enhance the credentials of GE.

C. General equilibrium theory and unemployment

Apart from consumption, GE theory also takes production into consideration.

To quote from Arrow's Nobel lecture again:

> "A productive unit or firm is characterized by a relation between possible outputs and inputs. A firm may have, of course, more than one output. Then firm f may be characterized by its transformation surface, defined by an equation, $T(y_{f1}, \dots, y_{fn}) = 0$, where y_{fi} is taken to be an output if positive and input if negative; the surface is taken to define the efficient possible input-output vectors for the firm, that is, those which yield maximum output of one commodity for given inputs and given outputs of other commodities. The optimizing behavior of the firm is taken to be the maximization of profit among the points on its transformation surface. Because of the sign conventions for inputs and outputs, the firm is seeking to maximize,

$$\sum_{i=1}^{n} p_i y_{fi} \qquad (4)$$

And a little later:

> "For any commodity i, there will be some demands and some supplies at any given set of prices. Following Hicks, we will speak of the *excess demand* for commodity i as the sum over all individuals and firms of demands and supplies, the latter being taken as negative. The demand by individual h is $x_{hi}(p_1, \dots, p_n)$, so that the total demand by all households is

$$\sum_{h} x_{hi}(p_1, \dots, p_n) \qquad (5)$$

> "The supply by households is the aggregate amount they have to begin with, i.e.,

$$\sum_{h} \bar{x}_{hi} \qquad (6)$$

> "Finally, the aggregate demand by firms is

$$\sum_{f} y_{fi}(p_1, \dots, p_n) \qquad (7)$$

> some firms may be demanders rather than suppliers, but the sign convention assures that the above sum gives the aggregate net supply by firms, i.e., after cancelling out demands by one firm which are supplied by another.

> "... Further, the satisfaction of the budget constraint for each individual also restricts the excess demand functions. Since for each individual, the monetary value of expenditure planned at any set of prices equals the monetary value of his initial endowments plus his share of the profits, we

have in the aggregate that the money value of planned expenditure by all households equals the money value of total endowments plus total profits, or

$$\sum_h \sum_{i=1}^{n} p_i x_{hi}(p_1, \dots, p_n) = \sum_h \sum_{i=1}^{n} p_i \bar{x}_{hi} + \sum_f \sum_{i=1}^{n} p_i y_i(p_1, \dots, p_n) \qquad (8)$$

or, from the definition of excess demand,

$$\sum p_i z_i(p_1, \dots, p_n) \equiv 0 \qquad (9)$$

[z_i is the market excess demand for commodity i] where the identity symbol reminds that this relation, called by Lange [1942] Walras' Law, holds for all values of the prices."

With those definitions in place and a long discursion through more math we come to the section "The Existence of Competitive Equilibrium"

> "A set of prices defines a competitive equilibrium if supply and demand balance on each market, including the possibility of corners, with *some* choice of the profit-maximizing input-output vector for each firm. Formally, we will say that a price vector p^*, an input-output vector y_f^* for each firm, and a consumption vector, $x_h^* = x_h(p^*)$, for each individual together constitute a competitive equilibrium if the following [four] conditions hold" (Arrow, 1972).

We are concerned here with the second condition.

> "(b) for each commodity i,
>
> $$\sum_h \bar{x}_{hi} + \sum_f y_{fi}^* \geq \sum_h x_{hi}^*" \qquad (10)$$

This is a straightforward conservation law. In plain English it says that the total demand of a commodity by households must be less than or equal to the sum of the initial holding of that commodity by households together with the output of that commodity by firms.

This applies to any commodity; therefore it must also apply to labour. Since households are suppliers of labour but not consumers of labour the right hand side of the above equation is equal to zero and the first term on the left hand side is positive. Similarly, since firms are consumers but not suppliers of labour the second term on the left hand side is negative.

So the equation above tells us that *the supply of labour will always be equal to or greater than the demand for labour*. This curious result derived from the equations of GE theory is in complete accord with reality (with the sign in practice being one of inequality) because we know that even in economies that are not undergoing a recession, there is a certain amount of unemployment, which is sought to be explained under the rubrics of frictional unemployment, non-accelerating inflation rate of unemployment, search-match delays and so on.

However, the equation allows a simpler, more logical explanation for such unemployment. Most households have only the single commodity called labour to sell. On the other hand, nearly all firms in all industries have to buy greater or lesser quantities of labour. Labour is the largest market in the economy. Compensation paid to employees in the US amounts to about 44% of GDP at present (St Louis Federal Reserve, 2016). So even when every market is only *slightly displaced from equilibrium*, to an extent that the disequilibrium is barely perceptible, the small displacements from perfect equilibrium in every market will be cumulative. That is, in the case of the input, labour, unemployment will amount to a relatively large proportion of the labour market, say, of the order of 5% in the US. The constant presence of unemployment is actually proof that the economy is always in the process of attaining equilibrium but never quite there. It is not credible to suggest this is entirely a matter of frictions, churning, and matching. There will always be more or less involuntary unemployment.

Unfortunately, this explanation is not permitted by the rules of GE theory, since equilibrium is attained.

To quote from Arrow's Nobel lecture again, the third condition for equilibrium is "(c) for any commodity for which the strict inequality holds in (b) [eqn 10], we must have $p_i^* = 0$" (Arrow, 1972)

Now this condition has been used earlier to account for the case of free goods. For example, the supply of air is greater than the demand for air, so its price must be zero. By that token, if the supply of labour is greater than the demand for air, its price must be zero.

We come away with the impression that GE theorists do not wish to listen to what their own equations are telling them but prefer to torture them so that the equations say what the theorists want to hear, viz. that the economy is in equilibrium.

D. The relationship between supply and demand

Again, in this section an extensive extract from Arrow's Nobel lecture is required. To get a proper grasp of the context it is advisable to read the complete speech, though it is not needed for our purpose.

> "We begin to see that a Pareto efficient allocation is an equilibrium of supply and demand in the generalized sense which includes corners. We also see that,
>
> $$\sum_{i=1}^{n} p_i(z_i - z_i^0) \geq 0 \quad for\ z\ in\ Z \qquad (11)$$
>
> [z_i is the excess demand vector for commodity i, Z is the set of all excess demand vectors and the superscript 0 refers to the Pareto-optimal case.]
>
> "Let us go back to the definition of excess demand, as a sum of individual and firm demands and supplies.

$$z_i = \sum_h x_{hi} - \sum_h \bar{x}_{hi} - \sum_f y_{fi} \qquad (12)$$

where $y_f = (y_{f1}, \dots, y_{fn})$ is a technologically possible vector of inputs and outputs for firm f and $x_h = (x_{h1}, \dots, x_{hn})$ is a possible vector of consumptions for individual h. In particular, the excess demands defined by the Pareto efficient allocation can be written in this form,

$$z_i^0 = \sum_h x_{hi}^0 - \sum_h \bar{x}_{hi} - \sum_f y_{fi}^0 \qquad (13)$$

and then, if z belongs to Z, we must have, for each h, that the consumption vector of individual h, $(x_{h1}, \dots, x_{hn})$ is preferred to that under the Pareto efficient allocation $(x_{h1}^0, \dots, x_{hn}^0)$. Then,

$$\sum_h \left(\sum_{i=1}^n p_i x_{hi} - \sum_{i=1}^n p_i x_{hi}^0\right) - \sum_f \left(\sum_{i=1}^n p_i y_{fi} - \sum_{i=1}^n p_i y_{fi}^0\right) \gg 0 \qquad (14)$$

if, for each h, x_h is preferred by individual h to x_{hi}^0.

"Now the elementary point about this inequality is that the variable vectors x_h, y_f are independent of each other. It is not hard to see that this inequality can hold only if it holds for each individual and each firm separately. For a firm f, this means that,

$$\sum_{i=1}^n p_i y_{fi}^0 \geq \sum_{i=1}^n p_i y_{fi} \text{ for all possible } y_f \qquad (15)$$

that is, if we interpret the $p_i's$ as prices, each firm is maximizing its profits. The corresponding interpretation for individuals is somewhat less simple; it is that the consumption vector prescribed by the given Pareto efficient allocation is the cheapest way of deriving that much satisfaction" (Arrow, 1972).

The details of the above derivation are not so important as the assumption that makes it possible:

> "Now the elementary point about this inequality is that the variable vectors x_h, y_f are independent of each other" (Arrow, 1972).

In Section A we established that GE theory and Marshallian demand analysis are equivalent. In this sense, GE does not represent an advance over Marshallian demand analysis. However, if we consider the assumption above, that consumption and production are independent of each other, it seems clear that GE introduces conditions that are a giant step backwards compared with Keynesian theory.

Indeed, I think I exaggerate only a little if I say that the central purpose of writing *The General Theory* was to show that consumption and production are not independent of each other as

classical economics had assumed. This would of course have been clearer if Keynes' legacy had not focused to such a great degree on the issue of investment.

It requires quite some flexing of the imagination to believe that the demand for bread is dependent on the price of steel, but it requires no imagination at all to perceive that the output of steel affects the demand for bread, by affecting the incomes paid out to labour involved in the production of steel. The demand for bread is a weak function of the price of steel. The demand for bread is also a weak function of the output of steel, but it is probably a stronger function than the first. So, the problems of GE theory can be traced to its fundamental assumptions.

GE theory assumes that the individual's demand for commodity i is $x_{hi}(p_1, \dots, p_n)$ whereas in reality it is $x_{hi}(p_1, \dots, p_n, y_{f1}, \dots, y_{fn})$. When aggregate income is constant the cross effects cancel out so that the first expression can be taken as a good approximation. This is because we are primarily interested in how demand *changes* in response to *changes* in prices, and vice versa. Similarly the firm's demand for commodity i in the general case is not $y_{fi}(p_1, \dots, p_n)$ but $y_{fi}(p_1, \dots, p_n, x_{h1}, \dots, x_{hn})$.

In order to solve for general equilibrium, using the simpler but erroneous expressions for consumer and firm demand, GE theory had to make a number of assumptions that all but removed any resemblance to reality. *This suggests that using correct but far more complex expressions for consumer and firm demand would make the system of equations unamenable to a solution.* This, of course is a typical problem with formalism as a constraint on how an economy is conceived, and continues to be a generalisable problem for many aspects of mainstream economics.

The fact that the consumer demand for a commodity is only a weak function of the output of other commodities means that it is only in the aggregate, and during recessions, that such effects are felt. And, as noted earlier, the effect of a fall in aggregate demand is first felt in the largest market of all, the labour market.

It must be here pointed out that the assumption that firms' profits are paid out to consumers *does not* integrate production and consumption. What matters is not profit but the amount paid out as income. In other words, what matters is not

$$\sum_f \sum_{i=1}^{n} p_i y_{fi} \tag{16}$$

where outputs are positive and inputs are negative, but the same quantity with no negative signs attached to inputs.

As a final issue, it seems worth emphasizing one of Keynes important insights that provides a useful contrast.[1] Keynes recognized that it was possible to solve this problem of inputs in the aggregate using income as an intermediate variable. A simple numerical example, with intertemporal substitution, using aggregate demand as the intermediating quantity, establishes this.

[1] In general it should be noted Keynes made a significant break with Marshall in so far as investment affects income paid out, in turn this affects consumption which affects revenues of firms, and so on…

Consider a simple, closed economy operating at full employment. It produces $90 worth of consumption goods. An equal amount is therefore paid out in income flows: wages, rent, interest, profit and so on. The saving rate is 10%, so $9 of these income flows is saved and $81 is spent on consumption goods. That means $9 worth of consumption goods is unsold. But then the economy also produces $10 worth of investment goods. An equal amount is therefore paid out as income. Since the saving rate is 10%, $1 of this is saved and $9 spent on consumption goods. Financial institutions of course turn the $10 saving into loans for investment. So all consumption goods are sold and savings are sufficient to pay for investment goods as in the table below.

	Consumption goods	**Investment goods**
Income	$90	$10
Consumption	$81	$9
Saving	$9	$1

Assume also that individuals have, on average, accumulated 20 years' worth of saving, which at present is valued at $200. Then thanks to a housing and stock market crash, an average of ten years' worth of accumulated saving is lost. This is not an outlandish figure; in the 2008 housing and stock market crash the median US household lost 18 years of real net worth (Federal Reserve, 2012) In an attempt to recover the lost net worth, households increase their saving rate from 10% to 15%. At this rate, they reckon they would take 20 years to recover their lost net worth. As a result of the higher saving rate, the money spent on consumption goods falls from $90 to $85. When manufacturers of consumption goods see their output remaining unsold they cut production to $85 and also cut purchases of investment goods to $8. Manufacturers of investment goods then cut production to $8. So total income now paid out is $93, a fall from $100 earlier. The economy now looks as in the table below.

	Consumption goods	**Investment goods**
Income	$85	$8
Consumption	$72.25	$6.8
Saving	$12.75	$1.2

Now total consumption is $79.05, so $5.95 worth of consumption goods remains unsold. As a result manufacturers of consumption goods cut their production further and curtail purchases of investment goods even more. The downward spiral continues unabated. It is of course possible to see that there is a damping factor. The saving in the first year is $13.95, so consumers might lower their saving rate a little as they begin to recover their lost net worth. But this must be balanced against the fact that consumers had expected a saving of $15, which did not materialize because the higher saving rate reduced the aggregate income. Whatever the details, the example illustrates the interconnection of production and

consumption, which Keynes of course recognized and which GE theory assumes out of existence.

If we refer back to the final extract from Arrow above it can be seen that the independence of production and consumption is a crucial assumption in proving profit maximization by firms and utility maximization by individuals. However, if the assumption is wrong, as we have shown it is, we can question both profit maximization and utility maximization. During a recession following a large asset market crash the minimization of consumption takes precedence over the maximization of utility. We suggest, therefore, that utility maximization is an idea applicable only to the special case of an economy in equilibrium, within a GE stylized mathematical construct (an extremely limiting situation and one in which aggregate demand is not falling). Similarly for firms during a recession, survival is more important than profit maximization.

Conclusion

In this short paper we have established that GE and Marshallian demand analysis are equivalent, and that GE theory is problematic in various ways. This has important implications for core Keynesian concepts, such as involuntary unemployment, as well as the continued relevance of his work. Furthermore, in so far as GE is the archetype of formalism, its limits demonstrate something about the limits of some uses of mathematics within the mainstream. This and other considerations regarding GE, are important for contemporary problems of New Classical Economics, DSGE and Real Business Cycle theory, though we do not discuss them in this paper.

References

Arrow, Kenneth (1972) 'General Economic Equilibrium: Purpose, Analytic Techniques, Collective Choice', Nobel Memorial Lecture, December 12,
http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/1972/arrow-lecture.pdf

Duffie, Darrell and Sonnenschein, Hugo (1989) 'Arrow and General Equilibrium Theory', *Journal of Economic Literature*, June
http://www.darrellduffie.com/uploads/pubs/DuffieSonnenschein1989.pdf

Keynes, J. M. (1936) *The General Theory of Employment Interest and Money*. London: MacMillan
St Louis Federal Reserve Bank (2016) Graph of ratio of labour compensation to GDP in the US, https://fred.stlouisfed.org/graph/?g=75e6

Federal Reserve (2012) 'Changes in U.S. Family Finances from 2007 to 2010: Evidence from the Survey of Consumer Finances;' *Federal Reserve Bulletin* 98 (2)
http://www.federalreserve.gov/pubs/bulletin/2012/pdf/scf12.pdf

Author contact: philip.george@gmail.com

SUGGESTED CITATION:
Philip George, "The mathematical equivalence of Marshallian analysis and 'general equilibrium' theory", *real-world economics review*, issue no. 77, 10 December 2016, pp. 73-84,
http://www.paecon.net/PAEReview/issue77/George77.pdf

real-world economics review, issue no. 77
subscribe for free

The trouble with production theory

Bernard C. Beaudreau [Université Laval, Québec, Canada]

Abstract
The edifice that is modern economics sits atop two pillars, namely consumer theory and producer theory. In turn, the latter sits atop what is perhaps the most important footing in all of economics namely production theory. After all, economics is the science of wealth, and since wealth is the result of material processes, it stands to reason that production theory stands at the very core of economics. Recently, "post-real" macroeconomics has been barraged with criticism (Romer 2016, Syll 2016), mostly directed at the "imaginary shocks" that lie at its core. In this paper, we argue that while "post-real" macroeconomics suffers from a number of shortcomings, the veritable problem is one that extends beyond macroeconomics to virtually all of economics, namely an archaic, simplistic, increasingly irrelevant, and scientifically misspecified approach to modeling wealth-producing material processes – in short, neoclassical production theory. It is shown that when more appropriate models of production are employed, most if not all the puzzles and paradoxes that plague economics disappear, as do imaginary shocks.

JEL Codes O40, O47, O57, Q43

Keywords production theory, process engineering, consilience

1. Introduction

The edifice that is modern economics sits atop two pillars, namely consumer theory and producer theory. In turn, the latter sits atop what is perhaps the most important footing in all of economics namely production theory. After all, economics is the science of wealth, and since wealth is the result of material processes, it stands to reason that production theory stands at the very core of economics. Recently, "post-real" macroeconomics has been barraged with criticism (Romer 2016, Syll 2016), mostly directed at the "imaginary shocks" that lie at its core.' In this paper, we argue that while "post-real" macroeconomics suffers from a number of shortcomings, the veritable problem is one that extends beyond macroeconomics to virtually all of economics, namely an archaic, simplistic, increasingly irrelevant, and scientifically misspecified approach to wealth-producing material processes – in short, neoclassical production theory.

Our basic argument is relatively simple and straightforward: neoclassical production theory is scientifically and empirically incorrect and a source of systemic irrelevancy. The evidence is there for all to see. Consider for example, the following facts.

i. Real business cycle analysis is based on imaginary shocks at a time when any and all shocks should be identifiable and quantifiable. After all, economics is not sub-atomic particle physics.
ii. Growth theory is also based on imaginary shocks, namely total factor productivity (TFP), which suffer from the same shortcomings.

iii. General Purpose Technology-based growth theory is based on shocks that while historical accurate have spawned the "electricity paradox" as well as the "information paradox."
iv. Neoclassical production theory violates the basic laws of mechanics according to which only force/energy can be physically productive. As labor is essentially supervisory in nature and capital is not a source of energy, it stands to reason that the notions of labor productivity and capital productivity sit in violation of basic physics.
v. Production theory is a misnomer, as what economics regards as theory is little more than simple and multiple correlations, with no underlying theory whatsoever.
vi. Material processes as modeled by economists and by process engineers are completely orthogonal. Notions such as the marginal product of labor, or capital have no equivalents in engineering or applied physics.

We will stop here. Our point is simple, namely that nothing can be right in economics if its most important cornerstone is not right. And that cornerstone is our understanding of the material processes that generate wealth.

This raises the obvious question of "how did this happen." Why is it that after over two centuries and a half of discourse, production remains the poor cousin of economics? How did a branch of moral philosophy which has sought and continues to seek knowledge and understanding degenerate into the realm of the metaphysical and imaginary, riddled with paradoxes and puzzles? After all, we are talking about material processes which are understood in all other fields of science (e.g. photosynthesis) with no exception.

In this paper, it will be argued that a combination of factors conspired to prevent economics from developing a scientifically accurate and relevant theory of production, consistent with the material process sciences in general. To illustrate these factors, the discussion will revolve around a comparison of the evolution of economics on the one hand, and thermodynamics on the other, two fields of scientific endeavor that were born of the steam engine and the ensuing industrial revolution (Berg 1980, Beaudreau 1998). We will show that repeated, unsuccessful attempts at closing the widening gap between the two left the door wide open for the metaphysical, the imaginary and the increasingly irrelevant. We also argue that this mattered little when growth was robust, but when the music died (i.e. when productivity growth slowed down), the profession found itself at a loss, one whose effects continue to be felt to this very day.

2. How we got here

In this section, we examine how production theory, over the years and decades, became increasingly irrelevant. This is accomplished by drawing a parallel between production theory in economics and the branch of physics known as thermodynamics, both of which were spurred on by the introduction of the Watts-Boulton rotary steam engine. It will be argued that despite a promising start, economics got side-tracked by non-production related issues, notably income distribution and the business cycle, while thermodynamics was free to evolve to the point of generating a set of laws that continue to hold today.[1] It will be shown that

[1] Lindenberger and Kummel (2002) make a similar argument, pointing out that: "the problem of the physical generation of wealth was coupled inseparably to the problem of the physical distribution of

despite a common starting point, the two diverged quickly with the result that by the end of the 19th century, production as formalized in economics was virtually orthogonal to production as formalized in applied physics and engineering. Interestingly, this duality found its way into the writings of a number of prominent figures in the 19th century.

Consider the following quotation, taken from Chapter 15 of Volume 1 of Karl Marx's Das Capital, entitled Machinery and Modern Industry.

> "Mathematicians and mechanicians, and in this they are followed by a few English economists, call a tool a simple machine, and a machine a complex tool. They see no essential difference between them, and even give the name of machine to the simple mechanical powers, the lever, the inclined plane, the screw, the wedge, etc. As a matter of fact, every machine is a combination of those simple powers, no matter how they may be disguised. From the economic standpoint this explanation is worth nothing, because the historical element is wanting. Another explanation of the difference between tool and machine is that in the case of a tool, man is the motive power, while the motive power of a machine is something different from man, as, for instance, an animal, water, wind, and so on. According to this, a plough drawn by oxen, which is a contrivance common to the most different epochs, would be a machine, while Claussens circular loom, which, worked by a single labourer, weaves 96,000 picks per minute, would be a mere tool. Nay, this very loom, though a tool when worked by hand, would, if worked by steam, be a machine. And since the application of animal power is one of man's earliest inventions, production by machinery would have preceded production by handicrafts. When in 1735, John Wyatt brought out his spinning machine, and began the industrial revolution of the 18th century, not a word did he say about an ass driving it instead of a man, and yet this part fell to the ass. He described it as a machine to spin without fingers.
>
> All fully developed machinery consists of three essentially different parts, the motor mechanism, the transmitting mechanism, and finally the tool or working machine. The motor mechanism is that which puts the whole in motion. It either generates its own motive power, like the steam-engine, the caloric engine, the electromagnetic machine, etc., or it receives its impulse from some already existing natural force, like the water-wheel from a head of water, the wind-mill from wind, etc. The transmit-ting mechanism, composed of fly-wheels, shafting, toothed wheels, pullies, straps, ropes, bands, pinions, and gearing of the most varied kinds, regulates the motion, changes its form where necessary, as for instance, from linear to circular, and divides and distributes it among the working machines. These two first parts of the whole mechanism are there, solely for putting the working machines in motion, by means of which motion the subject of labour is seized upon and modified as desired. The tool or working machine is that part of the machinery with which the industrial revolution of the 18th century started. And to this day it constantly serves as such a starting-point, whenever a handicraft, or a

wealth." The latter, we maintain, was in turn coupled inseparably to the problem of the business cycle. For example, Marx's interest in the question of distribution was motivated, in large measure, by his interest in the business cycle. This was also the case in the early 20th century where, for example, Paul Douglas' work on production was motivated ultimately by the business cycle.

> manufacture, is turned into an industry carried on by machinery" (Marx 1867, 261).

Clearly, Marx was familiar with both the principles of classical mechanics and the developing field of thermodynamics. Not only was he familiar with these branches of physics, but he ascribed much importance to motive power, to the steam engine, the caloric engine, and to the electromagnetic machine.[2] This stands in stark and obvious contrast with the first seven chapters of *Das Kapital* and the brunt of Marxist thought where the labor theory of value holds sway, notably that labor is the sole source of all value. In short, the labor theory of value is nothing more than classical production theory writ large. Hence, we have an example of the fundamental dichtomy that characterized the 19th century, namely the growing chasm and struggle between production as modeled in economics and production as modeled in the physical sciences.

Widening this chasm was a set of issues that were divorced from the question of production. Among these were (i) the business cycle and (ii) the question of income distribution. Without delving into the details, the labor theory of value was the key element of Marx's theory of surplus value and the ensuing policy prescriptions. A theory of value based on Chapter 15 of *Das Kapital* would have pre-empted all of Marx's substantive analysis, positive and normative. In fact, one could go as far as to argue that it would have pre-empted virtually all of radical economics.

Put differently, Marx's labor theory of value can be viewed as an intellectual ruse intended to justify a set of beliefs and principles regarding income distribution and the business cycle – specifically, the role of the rising rate of surplus value in precipitating the business cycle. He knew only too well that the steam engine had relegated labor to a supervisory role, and that material processes were, by then, powered primarily by motive power.

He was not alone. In the period before Marx (BM) and after Marx (PM), writers ignored the obvious, preferring to promulgate theories of material wealth that were increasingly orthogonal to classical mechanics and the developing field of thermodynamics. Consider the following cases. In the BM period, we consider the case of Adam Smith, who in the Wealth of Nations, attempted to analyze his friend Matthew Boulton's Watt steam engine. In the PM period, we consider the case of William Stanley Jevons who provided the classics' response (i.e. the establishment's response) to Marx's claim to the effect that capital was unproductive and hence profits and all forms of payments to capital were a form of theft.

2.1 Adam Smith and the Wealth of Nations

Before addressing the question proper, a few facts bear noting. First, Adam Smith, James Watt and Matthew Boulton were acquaintances. Second, James Watt had perfected his version of the Newcommen atmospheric steam engine in 1769, seven years before Smith published his magnum opus, the *Wealth of Nations*.

It is our view that the Wealth of Nations was, in large measure, a response to the steam engine and the enormous potential that it brought to bear on the British economy. And

[2] That he included the "electromagnetic machine" in 1867 should be seen as a testimony to his perspicacity and knowledge of the underlying engineering and physical principles of industrial material processes in the 19th century.

nowhere is this more evident than in Chapter 1 where he examined the question of labor productivity. In a nutshell, productivity had increased due (i) specialization (ii) lower downtime due to multiple tasks and (iii) machinery (Smith 1776). The latter were tools that had until then been powered by human muscle but were now powered by "fire power". Basing most of his observations on Matthew Boulton's Hockley Brook factory powered by Watt steam engines, Chapter 1 could well have been entitled "The Effects of Fire Power on Labor Productivity and Wealth."

This raises the question, why did Smith nest "fire power" in a discussion of labor productivity? After all, fire power had rendered labor's motive power (i.e. brawn) redundant. The reason, we believe, had to do with the politics of "fire power". Workers throughout Great Britain feared for their future as machines would replace artisanal weavers, spinners, metal workers, etc. By portraying machinery (fire power) as the friend and not the foe of conventional labor, he hoped to increase its acceptability.

The upshot, however, was disastrous for economics as it sired classical production theory where output was modeled as a function of labor and labor alone. Paradoxically, at a time when labor had been reduced to a supervisory role, it was elevated to the role of sole factor input – productive factor input, that is. However, in Smith's defense, the steam engine in the 1770s was as much an enigma to Newtonian physicists as it was to moral philosophers of the day.

2.2 William Stanley Jevons and the theory of political economy

Perhaps the most influential of 19th century iconoclasts – in large part, much in spite of himself, – was Willam Stanley Jevons, the father of neoclassical production theory. In the *The Theory of Political Economy* published in 1874, he outlined what was to become neoclassical production theory, namely that wealth is an increasing, continuous, twice-differentiable function of homogenous labor and capital. A lesser known, but equally important contribution, was his *The Coal Question An Inquiry Concerning the Progress of the Nation, and the Probable Exhaustion of Our Coal-Mines*, published in 1865 in which he addressed the question of Great Britain's dwindling coal reserves. In the opening salvo, he declared:

> "Day by day it becomes more evident that the Coal we happily possess in excellent quality and abundance is the mainspring of modern material civilization. As the source of fire, it is the source at once of mechanical motion and of chemical change. Accordingly it is the chief agent in almost every improvement or discovery in the arts which the present age brings forth. It is to us indispensable for domestic purposes, and it has of late years been found to yield a series of organic substances, which puzzle us by their complexity, please us by their beautiful colours, and serve us by their various utility.
>
> And as the source especially of steam and iron, coal is all powerful. This age has been called the Iron Age, and it is true that iron is the material of most great novelties. By its strength, endurance, and wide range of qualities, this metal is fitted to be the fulcrum and lever of great works, while steam is the motive power. But coal alone can command in sufficient abundance either the iron or the steam; and coal, therefore, commands this age – the Age of Coal.

> Coal in truth stands not beside but entirely above all other commodities. It is the material energy of the country – the universal aid – the factor in everything we do. With coal almost any feat is possible or easy; without it we are thrown back into the laborious poverty of early times" (Jevons 1865, 14).

Paradoxically, some nine years later (i.e. in 1874), coal or energy was completely absent from what is largely considered to be his magnum opus, namely *The Theory of Political Economy*, where capital is included in the production function and, more importantly, is assumed to be physically productive. In short, both labor and capital are assumed to be physically production and more importantly, are substitutable (Jevons 1874). One could argue that internal validity (i.e. *vis-a-vis* the debate over the role of capital in wealth) is what prevented Jevons from incorporating energy into the corpus of neoclassical analysis.

3. Growing dissidence

By the end of the 19th century, material processes as seen in economics and in thermodynamics were completely orthogonal, having little more than the dependent variable in common (i.e. output / wealth). Enter the second industrial revolution and the accompanying massive increase in energy consumption that resulted from the shift from shafting and belting to electric unit drive (Devine 1983, Sonenblum 1990), and you get the equivalence of an ecclessiastic philosophical crisis pitting engineers against economists. In this section, we examine the growing dissidence from outside – and to a certain degree, from within – of economics.

The second industrial revolution with its emphasis on a new power transmission technology that allowed for energy deepening – greater energy consumption per unit capital – brought out the glaring differences between economists and engineers over the underlying nature of material processes. As firms converted their cumbersome shafting and belting power drive technology to electric unit drive and cities, counties and states offered low-cost, grid-delivered electric power, productivity soared (Devine 1983, Sonenblum 1990, Beaudreau 1998). However, as neoclassical production theory ignored energy, these epoch-defining innovations failed to appear on the radar.[3] Moreover, as electric unit drive was less capital intensive than belting and shafting, investment expenditure actually fell as a result, making detection even less probable. .

The failure of the nation's leading economists to acknowledge the presence of what would be an epoch-defining change in material processes led to a chorus of criticisms, largely from statisticians and engineers. To them, the U.S. was in the midst of an industrial revolution, one to which the economics profession appeared to be oblivious. F.G. Tyron of the Institute of Economics (Brookings Institution) was among the first to point to the incongruity between production processes as modeled in economics and those observed in early 20th century America.

> "Anything as important in industrial life as power deserves more attention than it has yet received by economists. The industrial position of a nation may be gauged by its use of power. The great advance in material standards of

[3] In fact, it was only with the introduction of the notion of a general purpose technology (GPT) in the 1980s that the integral role of electricity in the second industrial revolution was formally acknowledged.

> life in the last century was made possible by an enormous increase in the consumption of energy, and the prospect of repeating the achievement in the next century turns perhaps more than on anything else on making energy cheaper and more abundant. A theory of production that will really explain how wealth is produced must analyze the contribution of this element of energy.
>
> These considerations have prompted the Institute of Economics to undertake a reconnaissance in the field of power as a factor of production. One of the first problems uncovered has been the need of a long-time index of power, comparable with the indices of employment, of the volume of production and trade, of monetary phenomena, that will trace the growth of the factor of power in our national development" (Tyron 1927, 281).

One year later (i.e. in 1928), Woodlief Thomas of the Division of Research and Statistics of the Federal Reserve Board published an article in the *American Economic Review* entitled "The Economic Significance of the Increased Efficiency of American Industry," in which he attributed the "striking changes" in American industry to power-related developments:

> "Large-scale production is dependent upon the machine process, and the increasing use of machinery and power and labor-saving devices has accompanied the growth in size of productive units. The growing use of power in manufacturing, for example, is reflected in the increase in horsepower of installed prime movers. This does not tell the whole story, moreover, for owing to increased use of electricity, the type of power used is now more efficient requiring less fuel and labor for its production. Out of a total installed horsepower in factories of thirty-six million in 1925, twenty-six million or 72 percent was transmitted to machines by means of electric motors, as compared with 55 percent in 1919, 30 per cent in 1909, and only 2 per cent in 1899. Between 1899 and 1925 horsepower per person employed in factories increased by 90 percent and horsepower per unit of product increased by 30 percent. Power has been substituted for labor not only through machines of production but also in the form of automatic conveying and loading devices" (Thomas 1928, 130).

In little time, this incongruity reached academia, specifically Columbia University where a group of engineers, known as the Technocacy Alliance outrightly rejected mainstream approaches to understanding wealth (essentially neoclassical production theory), arguing that they ignored mechanics, thermodynamics, process engineering and with the then state of the art regarding material processes in general.[4]

Foremost in the minds of the "dissidents" was the fact that while America's capacity to produce wealth was increasing, actual wealth appeared to be stagnant, prompting various calls to action. One such call came from the engineering department at Columbia, where Walter Rautenstrauch and Howard Scott launched the Technocracy movement. In short, it contended that mainstream economics in general and production theory in particular were irrelevant, not to mention incomplete and unscientific, and was in need of a major overhaul.

[4] While the Technocracy movement went through a number of iterations, organization-wise, our analysis will refer to the movement in general.

The latter would be grounded in thermodynamics in general and in energy in particular. In short, while not knowing it, the Technocrats were attempting to steer economics back on to a course similar to that taken by thermodynamics in the 19th century, one based on a the scientific underpinnings of material processes in economics.

Implicit in the Technocracy movement was a biting indictment of economics in general and production theory in particular. All material processes known to man were energy based, except in economics. There could be and would be no compromise. Neoclassical economics was rejected outright as scientifically irrelevant.

Driving the Technocracy movement was the view that energy-related innovations (electric unit drive in particular) had increased America's ability to produce without a concomitant increase in income and expenditure, leading to stagnation, unemployment and a full-blown depression. Technocracy offered both a diagnosis of the problem and a series of corrective measures/ reforms (an energy monetary standard, guaranteed income). The movement, however, lost much of its appeal with the rise of Keynesian economics, which provided a less radical fix. In short, animal spirits replaced the energy shock as the cause of the downturn.

Such boldness, especially from non-practitioners, was met with great resistance from the profession. For example, University of Chicago economics professor Aaron Director, in a pamphlet entitled, the Economics of Technocracy, seriously doubted its usefulness, arguing that mainstream economics and production theory was better suited to analyze the issues it sought to address. To begin, he summarized Technocracy in terms of six points:

i. The importance of energy: Through the expenditure of energy we convert all raw materials into products that we consume and through it operate all the equipment that we use." This, of course, has always been familiar to us, except that it was stated in terms of work, and not of energy. The great merit of the latter term is the possibility of dragging in the Law of Conservation of Energy and this marrying physics to the social mechanism.
ii. Energy can be measured, and the unit of measurement is always the same, while the dollar varies from time to time.
iii. The chief distinction between our society and that of all previous societies is the much greater amount of energy which can be generated. This has always been recognized by the designation of our civilization as the machine era.
iv. With every increase in the amount of mechanical energy the need for labor decreases.
v. The present depression marks the end of an era, since the increase in mechanical energy has at last become so great that, regardless of what happens, the need for human labor will rapidly decline.
vi. Does it follow, therefore, that the price system must break down, and that only he engineers can run a mechanical civilization (Director 1933, 8).

In short, according to Director, Technocracy offered nothing new, and, more importantly, was riddled with the most elementary of oversights and errors. Energy was nothing new, and, as such, presented no particular challenge to mainstream political economy. Technological progress, in this case, electric drive, increased, in a commensurate fashion, actual output and income, wages and profits. The causes of the Great Depression, he argued, lie elsewhere, notably in the war, the resulting debts, and tariffs without being more specific.

Another early 20th century dissenter was British Nobel-prize laureate chemist Frederick Soddy, who after his pioneering work with Ernest Rutherford on atomic transmutation turned his attention to economics, largely in response to the alleged "misspecification" of production theory, more to the point, to the absence of energy from the analysis. The gist of his critique can be found in the following allegory:

> "At the risk of being redundant, let me illustrate what I mean by the question, How do men live? by asking what makes a railway train go. In one sense or another the credit for the achievement may be claimed by the so-called engine-driver, the guard, the signalman, the manager, the capitalist, or share-holder, or, again, by the scientific pioneers who discovered the nature of fire, by the inventors who harnessed it, by Labour which built the railway and the train. The fact remains than all of them by their united efforts could not drive the train. The real engine-driver is the coal. So, in the present state of science, the answer to the question how men live, or how anything lives, or how inanimate nature lives, in the sense in which we speak of the life of a waterfall or of any other manifestation of continued liveliness, is, with few and unimportant exceptions, By sunshine. Switch off the sun and a world would result lifeless, not only in the sense of animate life, but also in respect of by far the greater part of the life of inanimate nature. The volcanoes, as now, might occasionally erupt, the tides would ebb and flow on an otherwise stagnant ocean, and the newly discovered phenomena of radioactivity would persist. But it is sunshine which provides the power not only of the winds and waters but also of every form of life yet known. The starting point of Cartesian economics is thus the well-known laws of the conservation and transformation of energy, usually referred to as the first and second laws of thermodynamics" (Soddy 1924, xi).

In short, according to Soddy, energy is the cornerstone of all human activity, including production. Labor, capital, information, technology etc. are all accessory inputs, necessary for but not the actual source of wealth. Despite much promise, the proposed Cartesian economics, based on the laws of basic physics (mechanics and thermodynamics) failed to make inroads into mainstream economics.

The fallout for engineering-based models of material processes is there for all to see. Energy would be absent from production theory, growth theory and economics in general until the 1970s. Interestingly, Edward Denison's encyclopedic view of the sources of growth (Denison 1962) ignored energy altogether, as did Zvi Griliches and other pre-1970 growth economists. There was, however, an exception, namely Roumanian economist Nicholas Georgescu-Roegen who in The Entropy Law and the Economic Process," provided an alternative account of production, one that he referred to as the flow-fund model of production. A "fund" factor is either labour power, farm land or man-made capital providing a useful service at any point in time; a '"stock" factor is a material or energy input that can be decumulated at will; and a "flow"' factor is a stock spread out over a period of time. The fund factors constitute the agents of the economic process, and the flow factors are used or acted upon by these agents.[5]

[5] Georgescu-Roegen contributed in a non-negligible way to the development of ecological economics where the economy is modeled in terms of entropy. While these models are consilient with basic physics and thermodynamics, they fail to address the issues raised by conventional economics, focusing for the

Despite its breadth, the "flow-fund model of production" suffered from a number of shortcomings. For example, the question of physical productivity was not adequately addressed as he implicitly assumed that capital and labor were physically productive. Further, both the breadth of his work (addressing such questions as sustainability) and the theoretical errors that were well documented, combined to cast a shadow on the flow-fund model as an accurate description of material processes, and on physics-based models of material processes. In other words, the controversy surrounding his use of entropy to describe not only energy but materials, and the myriad debates that ensued served to obfuscate his main contribution, namely that energy powered all known material processes.[6]

4. The straw that broke the camel's back: the productivity slowdown

Like a jilted lover, lady energy took her revenge in the 1970s with two crises, 1973 and 1979. Having ignored energy for the good part of two centuries, the profession now was faced with the task of understanding/analyzing the fallout. The problem was that production theory was devoid of energy. In fact, energy was no more than an intermediate input (U.S. Survey of Manufacturers), hence not productive.

Ernst Berndt and David Wood, in 1975, began to fill this void with the KLEMS production function, where energy, materials and services were added to capital and labor. The energy output elasticity was estimated at 4-6%, leading them to conclude that energy was a relatively unimportant factor input. The energy crises, they reasoned, would affect output via the respective cross-elasticities, specifically the energy-capital cross-elasticity.

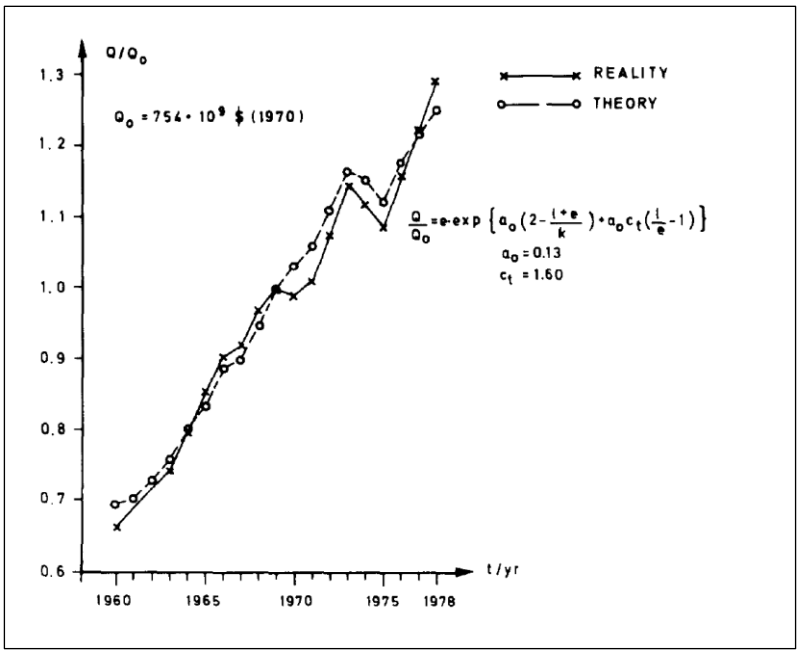
The Berndt-Wood estimates would go on to frame the debate over the role of the energy crises in the productivity slowdown. In short, most agreed that a quadrupling of the price of energy and the myriad energy shortages and embargoes were non-issues. There was, however, one exception, German physicist Reiner Kummel who argued the contrary, showing that productivity and growth tracked energy use/consumption (Kummel 1982). In short, the rate of growth of energy use/consumption had de facto slowed down, resulting in lower productivity and growth. Using an estimation technique known as LINEX, he reported an energy output elasticity of 0.58, significantly higher than that of Berndt and Wood (1975), one that was more in keeping with classical mechanics and thermodynamics. Beaudreau (1995) reported a similar elasticity of 0.53 for U.S. manufacturing (1950-1984), confirming Kummel's findings.

---

most part on sustainability and environmental degradation. Questions such as the role of labor and capital in material processes are not addressed, with the result that it is largely ignored in mainstream economics.

[6] Interestingly, Paul Samuelson described his work as perhaps the most important contribution of his era. Yet, energy and entropy continued to be absent from his work, acting as a metaphor for the role of entropy, namely important but ignored in formal analysis.

Figure 1 Actual and predicted growth of U.S. GDP 1960-1978



Source: Kummel (1982)

These findings did little to change the narrative. In short, energy continued to be regarded as a non-issue and an intermediate input. Moreover, production theory continued to be devoid of energy. Capital, labor and the Solow residual continued to garner virtually all of the attention.

Physics and the material sciences had made an incursion into economics via energy, but one that was short-lived and of no lasting consequence. Part of this had to do with the very manner in which it entered the equation, namely as an appendage to the traditional factors, capital and labor. In the KLEMS approach, capital, labor, energy, materials and services are all assumed to be physically productive. Even Kummel, who is a theoretical physicist, viewed capital and labor as physically productive, in violation of the laws of classical mechanics.[7] As Marx and Soddy had pointed out, motive power drives modern material processes, with labor and capital as passive (i.e. non-physically productive) inputs.

In the next section, we present the few attempts were made at consilient economics-material sciences approaches to modeling production. That is, models that are consistent with basic physics and address the issues raised by economists. It bears reminding that while both process engineers and economists study material processes, they do so for different reasons. The latter do so to better understand the wealth creating process, while the former do so to improve efficiency, or solve bottlenecks.

[7] Kummel's LINEX production function is typically specified as follows: $q_{CDE} = q_0 k^{\alpha_0} l^{\beta_0} e^{\gamma_0}$, where k is the capital input, l, the labor input, and e, the energy input.

5. Consilient models of production

As evidenced by the perspicacity of his insights into the fundamental nature of material processes in the industrial era, Karl Marx would have to be considered one of the earliest writers to present a "consilient" account/model of the wealth-creating process. Others include the Technocrats, with their emphasis on energy and its primordial role in material processes. However, neither Marx nor the Technocrats offered an analytical model of material processes, comparable in nature to the classical or neoclassical production function.

$$W(t) = n[T(t), S(t), I(t)]E(t) \quad (1)$$

Beaudreau (1998) presented the consilient energy-organization approach to modeling material processes, where he identified to universal factor inputs, namely broadly-defined energy and broadly-defined organization. Drawing from process engineering, he argued that the former was physically productive while the latter was organizationally productive.[8] Equation 1 presents the gist of the model, where W (t) is output, E(t) is energy consumption and η is second-law efficiency. The latter, in turn, is a function of S(t) the supervisory input, T (t), tools, and I(t) information. In keeping with basic physics, the latter three factor inputs are not physically productive, but rather are organizationally productive affecting second-law efficiency. Better tools (i.e. James Watt's external condenser, electric unit drive) increase energy efficiency by minimizing losses. As η is bounded from above, it stands to reason that organizational innovations will have limited effect on output and output growth.

Using data from U.S., German and Japanese manufacturing, he showed, like Kummel (1982), that productivity and growth tracked energy consumption, and more importantly, that the productivity slowdown coincided with a marked decrease in energy consumption growth, which in turn, coincided with the energy crisis. Table 1 reports the relevant growth rates for manufacturing value added Q, electric power EP, labor L and capital K, as well as the relevant fixed-weight aggregate input index (AI) for three time intervals: 1950–1984, 1950–1973, and 1974–1984. We see that virtually all growth in output in the pre- and post-productivity slowdown periods (i.e. before and after1973) is accounted for by growth in the inputs, with energy being the main driver.

In addition to rationalizing the productivity slowdown in terms of energy consumption growth, the energy-organization approach offers a simple, yet compelling account of the "information paradox," namely that information is a non-physically productive organizational factor input whose effect on second-law efficiency was not only punctual in nature, but theoretically limited – in the limit, zero.[9] In other words, because information is a non-physically productive input, it follows that the information paradox is not a paradox at all, but rather a manifestation of this simple fact.

[8] Process engineers typically consider two inputs, namely energy and information. See Alting (1994).
[9] This corresponds to the case in which second-law efficiency has reached its theoretical maximum, leaving little room for further information-based improvements. See Beaudreau and Lightfoot (2015) for further details.

Table 1: Output and input growth rates: U.S., German and Japanese manufacturing

US			
	1950-1984	**1950–1973**	**1974–1984**
USV A	2.684	3.469	0.121
USAI*	2.674	3.472	0.310
USEP	4.052	5.371	0.246
USN	0.662	0.900	-0.091
USK	3.694	3.614	3.4008
Germany			
	1963–1988	**1962–1973**	**1974–1988**
GERV A	2.462	6.522	1.486
GERAI*	2.433	5.190	1.080
GEREP	2.894	5.883	1.366
GERN	-0.785	0.592	-0.938
GERK	2.945	5.620	1.406
Japan			
	1965–1988	**1965–1973**	**1974–1988**
JAP V A	3.826	8.844	3.099
JAP AI*	3.566	9.856	1.538
JAPEP	3.559	11.320	0.965
JAPN	-0.082	2.297	-0.367
JAPK	7.520	13.536	5.182

* $\hat{\beta}_1 \frac{\dot{ep}(t)}{ep(t)} + \hat{\beta}_2 \frac{\dot{l}(t)}{l(t)} + \hat{\beta}_3 \frac{\dot{k}(t)}{k(t)}$, where $\hat{\beta}_i$ the estimated output elasticities, are taken from Beaudreau (1995).

Source: Beaudreau (1998).

6. The potential fallout from abandoning poor science by field

Consilent approaches to understanding material processes have important implications not only for growth, but for virtually all of the current fields and sub-fields of economics. This owes, in large measure, to the key role wealth and material processes play in each. In this section, we examine three such fields, namely labor economics, capital theory and finance and lastly, income distribution.

6.1 Labor economics

The role of what is commonly referred to as the labor input in modern (read: industrial) material processes has been the subject of vigorous debate from the very beginning. After all, the Watt-Boulton rotary steam engine had rendered human force/brawn redundant, relegating labor to a supervisory role, one that could be easily accomplished by women and even

children. It was clear to most writers that labor was no longer physically productive, but organizationally necessary as machines were prone to stoppages and breakdowns.

As it no longer supplied the force driving tools (simple and complex), its claim on the spoils was increasingly under siege. Theoretically, there was no reason it should benefit from the resulting energy deepening (i.e. greater use of force) and greater productivity. From a strict "productivity" point of view, its contribution had been diminished, which would/could justify a lower wage.

Classical and neoclassical production theory, however, continued to put labor at the center of its analyses of material processes. The latter went as far as to attribute three-quarters of output to what had become a supervisory input – what Alfred Marshall referred to as "machine operatives."[10] Physically unproductive tools and equipment (i.e. capital) were assigned a secondary role, contributing to one-quarter of output. Ironically, energy, the only truly physically productive input, was ignored.

This brings us to consilient models of material processes and their implications for labor economics. The first implication is definitional and descriptive, namely accurately describing the role of modern-day labor in production processes. In short, labor is a supervisory input, and one whose role in production is currently under siege from ICT-based control devices. In other words, Marshall's machine operatives are being replaced with automated control systems. In time, there is reason to believe that labor as we know it today will disappear, bringing with it, the end of human supervision. In his 1995 best-selling book, the End of Work, Jeremy Rifkin makes a similar argument. However, it should be pointed out that his choice of title is misleading as work will continue unfettered, it being the product of the continuous application of force.[11] A more appropriate title would have been The End of Human Supervision of Material Processes.

6.2 Capital theory and finance

The role of tools (simple and complex) in material processes has not only been the subject of great debate, it has spawned some of the most notable schisms in economics. Is capital productive, and if so, what is the extent of its contribution? The classics viewed it as a complement to labor, but not productive in and of itself – consistent with classical mechanics. As pointed out above, this led to the first schism in economics when Karl Marx declared, on the basis of classical production theory, that profits were a form of theft, given that capital was not physically productive. To remedy the situation and restore legitimacy to capitalism, the neoclassical writers simply decreed it to be physically productive, thus entitled to a share of the output.

Consilient approaches to production invariably invoke the laws of classical mechanics with regard to capital. Put differently, tools and equipment are not, per se, physically productive, providing mechanical advantage.[12] This has important implications for capital theory and

[10] Interestingly, the British from Marshall on, referred to labor as "operatives," and not workers, the implication being that labor "operated" machinery, and did not "work" in the traditional sense.
[11] As odd as it may sound, workers stopped working with the introduction of inanimate power / energy / force, metamorphisizing into what was a supervisory input.
[12] Mechanical advantage is a measure of the force amplification achieved by using a tool, mechanical device or machine system. Ideally, the device preserves the input power and simply trades off forces against movement to obtain a desired amplification in the output force. The model for this is the law of

finance as both are based on the notion/belief that capital is productive and hence can legitimately lay claim to a share of the spoils. In its absence, the theory breaks down, raising the question how should the owners of tools and equipment be remunerated? What share of output should they receive? One thing however is obvious, namely that they, no more than labor, are entitled to energy's productivity.

6.3 Income distribution

As can be surmised, the implications of consilient models of material processes for the distribution of income are far-reaching. The neoclassical approach based on labor and capital physical productivity is rendered null and void as neither of these two factor inputs is physically productive. While the question of alternatives is beyond the scope of this paper, we can nonetheless speculate as to possible replacements.

Were we to invoke a pure productivity standard, it is clear that all output would revert to the owners of the sole physically productive factor, namely energy. However, such a standard would not be incentive compatible for labor and capital. Another possibility would be to provide labor and capital with their reservation factor payment so as to entice participation. Put differently, the owners of energy would share in the spoils in order to ensure that organization be present. A third possibility would be for the owners of energy to share what Beaudreau (1998) referred to as "energy rents," defined as the difference between the value of energy's physical product and its cost.[13] Accordingly, the greater part of the wages that are now paid to labor would consist of energy rents, as would the greater part of profits paid to capital. Such a view is implicit in David Lloyd George's view of wages noted in Coal and Power in 1924:

> "Those people are best paid and most prosperous that make use of the resources of science.....the average level of earnings must depend on production and production increases as the use of power per head of population increases" (Lloyd George 1924).

As such, variations in income shares over time (Piketty 2014) can be seen as resulting from changes in the bargaining game over energy rents. As Piketty has shown, capital's share increased in the 1920s and from the 1980s onwards. The former can be attributed to the energy deepening in the form of electric power in the 1920s, while the latter can be attributed to ICT-based factory automation, both of which allowed the owners of capital to appropriate a larger share of the associated energy rents.

---

the lever. Machine components designed to manage forces and movement in this way are called mechanisms. An ideal mechanism transmits power without adding to or subtracting from it. This means the ideal mechanism does not include a power source, is frictionless, and is constructed from rigid bodies that do not deflect or wear. The performance of a real system relative to this ideal is expressed in terms of efficiency factors that take into account friction, deformation and wear (https://en.wikipedia.org/wiki/Mechanicaladvantage).

[13] 13Lindenberger and Kummel (2002) made a similar argument in what is a different context. Specifically, they pointed out that: "If wealth had been distributed according to the "marginal productivity theory", labor would have received only a share of national income much smaller than the observed 70%. But apparently, in the past most of the value added by energy was attributed to labor. The underlying mechanism of distribution was that of wage-negotiations in which free labor unions, powerful during times of high employment, regularly succeeded in winning wage increases according to the growth of productivity, i.e. increased production due to increased and more efficient energy utilization. This way most of the population in the industrialized countries benefited from the wealth generated by the production factors capital, labor, and energy.

7. The fault, dear Paul, lies not in our stars, but in ourselves, that we are underlings

In his analysis of post-real macroeconomics, Paul Romer raised an extremely important question, namely why has poor science (i.e. imaginary shocks) been tolerated (Romer 2016)? The main reason, he argued, lies with the sociology of science, with the presence of steep dominance hierarchies in macroeconomics, controlled by a few gatekeepers or kingpins.

While we believe that this is part of the reason, we believe that it is incomplete. We believe that the principal reason the profession has tolerated (in the sense of not rejecting) and continues to tolerate and reinforce constructs like imaginary shocks has to do with the current state of production theory where residuals have become the "acceptable" norm. In short, it has to do with the primitive yet universally accepted accounts of production, accounts that continue to be orthogonal to the material sciences in general.

As the Solow residual is universally accepted, the notion that it varies over time, and in some cases, takes on negative values, is not as much of a stretch as one would think. In other words, existence and variability conspired to provide legitimacy to the notion of a systemic "negative" productivity shock. The "kingpins" or "gate keepers" got away with it because they appealed to something that is second nature to the current generation of economists, namely the Solow residual, and variability over the spectrum of real numbers. Today, the Solow residual is as much a part of growth theory and economic theory in general as are the traditional factor inputs, capital and labor. Put differently, Kydland and Prescott would not and could not have gotten away with "imaginary shocks" had they not been part of the intellectual landscape that is present day economics.

8. Conclusions

It is our view that until the economics profession gets production right, nothing can be and will be right in economics. For over two centuries, we have failed to heed the advice of physicists, production engineers, engineers in general, even chemists, preferring to promulgate models that while convenient and self-serving, are nonetheless theoretically misspecified and stand in violation of the basic laws of physics.

Economics is the only field within the material sciences in which growth is still shrouded in mystery. We submit that the reason has to do with the lack of consilience or pluralism. Instead of rendering our models consistent with those used elsewhere in the material sciences, we have created a world onto ourselves in which the laws of physics are suspended or even worst, violated. It therefore comes as little surprise that applied fields like macro or economic growth rely on imaginary concepts.

It is our view that a revolution of sorts is required in the field of production theory. However, we hasten to add that in actual fact it would be less of a revolution, and more of an acknowledgement of over a century of parallel thinking and theorizing. While the fallout will be great for labor economics, capital theory and finance and income distribution, we feel that we have no choice. Either we continue to be irrelevant, or we begin the long and arduous task of rendering our formalizations of material processes consistent with those found in the material sciences (i.e. consistent with basic physics), and in so doing, of putting economics on the track of scientific legitimacy and respectability.

It is our view that had economics heeded the advice of the dissenters in the 1930s, it would have sparred itself much turmoil, including post-WWII growth theory based on the imaginary Solow residual, post-productivity slowdown RBC macroeconomics, also based on the imaginary Solow residual, and 30 years of post-productivity growth theory (Romer, Lucas, Aghion and Howitt) that has failed to understand these very shocks. Further, we feel that until we acknowledge the basic fact that material processes in economics are analogous to those found in the material sciences in general, and we incorporate this fact in the modeling of production pro-cesses, economics will continue to be stricken with paradoxes, puzzles and imaginary constructs, not to mention irrelevancy.

References

Alting, Leo (1994) *Manufacturing Engineering Processes*. New York, NY: Marcel Decker Inc.

Beaudreau, Bernard C. (1995) "The Impact of Electric Power on Productivity: The Case of U.S. Manufacturing 1958–1984," *Energy Economics*, 17(3), 231–236.

Beaudreau, Bernard C. (1998) *Energy and Organization: Growth and Distribution Reexamined*. Westport, CT: Greenwood Press.

Beaudreau, Bernard C. and Douglas Lightfoot (2015) "The Physical Limits to Economic Growth by R and D-Funded Innovation," Energy, 84.

Beiser, Arthur (1983) *Modern Technical Physics*. Menlo Park, CA: The Benjamin/Cummings Publishing Company.

Berg, Maxime (1980) *The Machinery Question and the Making of Political Economy*. Cambridge: Cambridge University Press.

Berndt, Ernst and David Wood (1975) "Technology, Prices and the Derived Demand for Energy," *Review of Economics and Statistics*, 57(3), 259–268.

Denison, Edward F. (1962) *The Sources of Economic Growth in the United States and the Alternatives Before Us*. New York, NY: Committee for Economic Development.

Denison, Edward F. (1985) *Trends in American Economic Growth 1929–1982*. Washington, DC: The Brookings Institution.

Devine, Warren D. (1990) "Electricity in Information Management: The Evolution of Electronic Control," in Schurr, Sam. H. et al. (eds.), *Electricity in the American Economy*. Westport CT: Greenwood Press.

Director, Aaron (1933) *The Economics of Technocracy*. Chicago, IL: University of Chicago Press.

Georgescu-Roegen, Nicholas (1971) *The Entropy Law and the Economic Process*. Cambridge, MA: Harvard University Press.

Griliches, Zvi (1994) "Productivity, R and D, and the Data Constraint," *American Economic Review*, 84(1), 1-23.

Jevons, W.S. (1865) *The Coal Question, Chapter 1*. London: MacMillan and Co.,

Jevons, W.S. ([1874]1965) The Theory of Political Economy, Augustus M. Kelley, Bookseller, New York.

Kummel, Reiner (1982) "The impact of energy on industrial growth," *Energy* 7 (2): 189-201

Kummel, Reiner, Dietmar Lindenberger, and Wolfgang Eichhorn (1998) "The Productive Power of Energy and Economic Evolution," *Indian Journal of Applied Economics, Special Issue on Macro and Micro Economics*.

Dietmar Lindenberger and Reiner Kummel, (2002) "Thermodynamics and Economics," *Post-Autistic Economics Review*, 14, June 24.

Lloyd-George David (1924) *Coal and Power*. London: Hodder and Stoughton.

Marx, Karl, (1906) "Machinery and Modern Industry," in Marx, Karl, *Das Kapital* (Chapter 15), New York, NY: The Modern Library.

Piketty, Thomas (2014) *Capital in the 21st Century*. Cambridge, MA: Harvard University Press.

Rifkin, Jeremy (1995) *The End of Work*. New York, NY: Putnam Publishing Group.

Romer, Paul (2016) *The Trouble With Macroeconomics*, Stern School of Business, New York University. https://www.law.yale.edu/system/files/area/workshop/leo/leo16romer.pdf

Rosenberg, Nathan (1972) *Technology and American Economic Growth*. Armonk, NY: M.E. Sharpe.

Rosenberg, N. (1983) "The Effects of Energy Supply Characteristics on Technology and Economic Growth," in Schurr, S. et al. (eds.) *Energy, Productivity and Economic Growth*. Cambridge, MA: Oelgeschager, Gunn and Hain.

Smith, Adam (1990) "Of the Division of Labour," (Book I, Chapter 1), *An Inquiry into the Nature and Causes of the Wealth of Nations*, Chicago, IL: Encyclopaedia Britannica, 1–12.

Soddy, Frederick (1924) *Cartesian Economics, The Bearing of Physical Sciences upon State Stewardship.* London: Hendersons.

Sonenblum, Sidney (1990) "Electrification and Productivity Growth in Manufacturing," in Sam H. Schurr, Calvin Briwell, Warren D. Devine Jr. and Sidney Sonenblum, eds. *Electricity in the American Economy: Agent of Technological Progress*. Westport, CT: Greenwood Press.

Syll, Lars Palsson (2016) "Paul Romers assault on post-real macroeconomics", *Real-World Economics Review*, 76, 30 September, pp. 43-54. http://www.paecon.net/PAEReview/issue76/Syll76.pdf

Thomas, Woodlief (1928) "The Economic Significance of the Increased Efficiency of American Industry," *American Economic Review*, 18, 122–138.

Tyron, F.G. (1927) "An Index of Consumption of Fuels and Water Power," *Journal of the American Statistical Association*, 22, 271–282.

U.S. Department of Labor, (1967) *Productivity and Technological Change in the Petroleum Refining Industry.* Department of Labor Statistics, Washington, DC.

Author contact: bernard.beaudreau@ecn.ulaval.ca

SUGGESTED CITATION:
Bernard C. Beaudreau, "The trouble with production theory", *real-world economics review*, issue no. 77, 10 December 2016, pp. 85-102,
http://www.paecon.net/PAEReview/issue77/Beaudreau77.pdf

real-world economics review, issue no. 77
subscribe for free

Trade surpluses of countries trading with the United States

John B. Benedetto [USA][1]

Abstract
This paper briefly summarizes trends in recent U.S. trade deficits, and places them in the context of how economists think about trade balances and their implications for national economies. It then looks in more detail at the inverse of U.S. trade deficits, i.e., the trade surpluses that U.S. trading partners have been running with the United States. By presenting basic data from the U.S. and other governments, it shows that many large U.S. trading partners have been running trade surpluses that account for a large portion of their GDP every year for many years. The paper concludes that, from an initial look at basic government data, reducing the U.S. trade deficit could mean a large adjustment not only for the United States, but also for many U.S. trading partners.

Introduction

The U.S. trade deficit has been persistently over two, and mostly over three, percent of U.S. GDP for several decades. When the United States (or any country) runs a trade deficit, its trading partners by definition run a trade surplus with the United States (or that country). While some Keynesian and heterodox economists have expressed concerns about the size and persistence of the U.S. trade deficit, the modern U.S. economics profession has been mostly dismissive of such concerns.

Arguably, modern U.S. economists have engaged in even less analysis of the effect of the U.S. trade deficit on its trading partners. This paper describes what basic data say about the scale of that effect, within the context of a brief summary of how economists think about trade balances. Hopefully, such a description can be a starting point for understanding not only the potential cost (or benefit) to the United States of persistent trade deficits, but which countries are potentially benefitting (or losing) and by how much.

This paper will first summarize briefly how economists think about trade balances. Second, it will offer a brief history of the U.S. trade deficit. Third, it will proceed to present basic data on the scale of the U.S. trade deficit relative to GDP growth in the countries with which the United States runs the largest trade deficits. Some further analysis follows.

How economists think about trade balances

Most economists use gross domestic product (GDP) as a measure of a country's economic activity within a period of time (usually a year or a quarter). GDP is defined as a country's

[1] Economist with the U.S. International Trade Commission. The views and conclusions expressed in this article are solely those of the author, in his personal capacity. They are not necessarily the views of the United States, the U.S. International Trade Commission, or any individual Commissioner. The author thanks Wendy Willis for her contributions. All errors are the author's. Please direct all correspondence to John Benedetto.

consumption plus its investment plus its government spending plus its trade balance (a surplus if the country's exports are more than its imports, and a deficit if imports are more than exports). That is, the classic formulation is:

GDP = Consumption + Investment + Government spending + Exports minus Imports

As a matter of arithmetic, a trade surplus (that is, exports being greater than imports) adds to a country's GDP, while a trade deficit (exports being less than imports) subtracts from GDP.

For many Keynesian economists, trade deficits should be avoided when a country and/or its trading partners are not already at full capacity utilization, because trade deficits mean lower national income. Closing a trade deficit when the economy is at less-than-full capacity utilization could boost a country's GDP without offsetting side effects, similar to the arithmetic noted above.[2] Keynesians have been making such arguments since the time of John Maynard Keynes himself, and the arguments enjoyed a particular strength from the 1950s through the 1970s, when numerous well-known economists encouraged countries to reduce their trade deficits to spur economic growth and increase employment. Keynes himself argued in multiple fora that closing trade deficits was important to eliminate unemployment and have international economic stability.[3] More recently, prominent economist Larry Summers may have rediscovered some of these arguments.[4]

On the other hand, for the most part, many modern (and especially U.S.) economists have argued that trade deficits should not be a concern for U.S. (or sometimes any country's) policymakers. These economists make a variety of arguments for why one should not necessarily be concerned about U.S. (or other) trade deficits. Some describe a country's trade deficit as simply the result of a surplus of foreign investment over domestic savings, and not a problem. For these economists, a trade deficit can provide other benefits to the trade deficit country, benefits such as increased investment in domestic industries, lower prices to consumers, and increased competitiveness of domestic industries.[5] They may argue that if the U.S. trade deficit were to go away, then U.S. GDP might suffer in other areas, such as from lower consumption or lower investment. They may also argue that closing a trade deficit could reduce employment, or that a trade deficit may reflect foreigners' sense that investments in the United States are secure.[6]

---

[2] For example, economist Dean Baker, discussing the U.S. trade deficit in 2014, stated "We must either have more consumption, more investment, or more government spending, or some combination, than would normally be the case because we have a trade deficit of 3.0 percent of GDP. Again, this is accounting identity stuff, it has to be true. If you disagree, read this as many times as necessary until you understand." See Baker, "'Secular Stagnation,'" November 2014.

[3] See discussions of Keynes' views, and the views of his followers, including Kaldor and Robinson, in Milberg, "Say's law in the open economy," 2002, pp 240-245; Lavoie, *Post-Keynesian Economics*, 2014, pp. 492-511; and Irwin *Against the Tide*, 1996, pp. 200-202. The word "Keynesian" is difficult to define exactly, as some have argued that Keynes' views are more radical than some of his followers, both in the immediate decades after Keynes, and especially since the advent of New Keynesians. This paper uses the word "Keynesian" to mean more generally the views of Keynes and his more immediate followers, and not the modern New Keynesians. See N. Smith "How 'Keynes' Became a Dirty Word," September 2014, Y. Smith, *Econned*, 2010, pp. 35-42, and Keen, *Debunking Economics*, pp. 199-210.

[4] See Ip, "The Case for Free Trade is Weaker Than You Think," Wall Street Journal, April 2016.

[5] See examples of these arguments listed in Benedetto, "Who Financed Recent U.S. Trade Deficits?" May 2014.

[6] See, for example, Perry "Imports and trade deficits" May 6, 2015 and Scissors "The trade deficit" March 16, 2015.

Alternatively, other modern economists describe trade deficits (as part of a current account deficit) as the sign of a lack of national savings, i.e. a reflection of other, "macro" economic problems, rather than a problem in and of itself.[7] This argument is fundamentally different than the above arguments, in that it is acknowledging that the trade deficit is associated with negative economic effects, instead of claiming it is actually a positive good for the trade deficit country. However, it is blaming the trade deficit on problems not directly related to trade, and thus alleging that it cannot be solved through trade policies.

As noted earlier, modern economists' lack of concern over trade deficits is relatively new (or a revivification of older ideas), and a shift from the Keynesian view that held more sway before the 1970s. Economist Douglas Irwin attributes the decline in modern economists' concern about trade deficits to their faith in the argument that flexible exchange rates will balance trade.[8] On the other hand, economists Dean Baker and Jared Bernstein explain that differences in economists' approaches toward trade deficits are in large part explained by whether a country is at full capacity utilization or not. At full capacity utilization, a trade deficit provides less expensive products for domestic consumers. At less-than-full capacity utilization, Keynesian economists emphasize that a trade deficit is a drain on jobs and production, and hence national growth.[9]

This paper is not analyzing which of these arguments is correct. However, it attempts to add to the debate by summarizing the other side of the historical U.S. trade deficit ledger, i.e., the effect of the U.S. trade deficit on U.S. trading partners. Using the logic of modern economics, U.S. trading partners are hurting themselves by running such trade surpluses, and would benefit from reducing them, although there may be some adjustments. Using Keynesian logic, though, these countries may benefit from these surpluses, and would thus likely be loath to let them go.[10]

Understanding the relative scale of other nations' trade surpluses with the United States can also help answer some general questions, no matter what one's analysis of the effects of U.S. trade deficits. For example, in 2005, then-U.S. Federal Reserve Chairman Ben Bernanke famously called for rebalancing global trade and investment accounts.[11] Nonetheless, Bernanke's analysis has been criticized by some Keynesian and heterodox economists, even

[7] See Stein, "Balance of Payments," 2008 and Benedetto "Who Financed Recent U.S. Trade Deficits?" May 2014, for a summary of some of these arguments, including those which focus on capital flows or the effects of having a reserve currency. Milberg also shows the parallel between the arguments of those economists who argued with Keynes in the 1930s and modern economist Paul Krugman's denunciations of those concerned with balanced trade in the 1990s. See Milberg, "Say's law in the open economy," 2002, pp. 251-252.
[8] Irwin, *Against the Tide*, 1996, pp. 202-203.
[9] See Baker, "'Secular Stagnation,'" November 2014; Bivens, "Yes, Trade Deficits," May 2015; and Bernstein, "What we're arguing about ," May 2015, with Bernstein citing further work by Federal Reserve Chairman Ben Bernanke and economist Larry Summers. This difference between economists who worry about the trade deficit and those who do not has been widely recognized as the root of economists' debates over trade deficits for a long time. For example, economist Douglas Irwin cites famed economist J.R. Hicks as having been persuaded by Keynes' arguments that the classical arguments for free trade were not completely correct. See Irwin, *Against the Tide*, 1996, pp. 202-204. More recent Keynesians, such as Thomas Palley, posit that demand deficiencies in emerging economies lead some governments of emerging economies to engage in currency and other policies to maintain trade surpluses with the United States. See Palley, "The theory of global imbalances," 2014, at pp. 23-24.
[10] See, for example, Schumacher "Deconstructing the Theory of Comparative Advantage" p. 96 for additional explanation of why, when the assumptions used by modern trade economists are relaxed, countries benefit from trade surpluses.
[11] Bernanke, "Remarks," March 2005.

if they also in the end call for policies that would close U.S. trade deficits.[12] While not taking a position on this debate, the data in this paper give a rough idea of how much trade balance adjustment such proposals would require of U.S. trading partners.[13] Additionally, whether one regards the U.S. trade deficit as a benefit or a cost to the U.S. economy, understanding its scale compared to other countries' economies will help in understanding from which countries this benefit or cost comes.

A brief history of the U.S. trade balance

Historically, the United States had roughly balanced trade from the 1790s until the 1970s.[14] In 1861, when Southern legislators left Washington, the remaining Congress and President Buchanan passed high tariff increases, as advocated by the incoming President Lincoln. Tariff levels then remained high for most of the period after 1860. During this period, while the small trade deficits of the pre-Civil War period were replaced by small trade surpluses, the United States remained a nation for which trade was a small share of GDP, and trade was roughly in balance. Despite some claims,[15] the United States clearly did not develop behind large mercantilist trade surpluses, as no large surpluses ever existed outside a very brief period after World War II. [16]

After World War II, the United States ran large trade surpluses with the rest of the world, from 1942-1952. However, after this brief period, U.S. trade balances went back to their traditional small surpluses. For almost 30 years after World War II, trade was not a large share of U.S. GDP, and U.S. trade was roughly in balance.[17]

By the 1970s, U.S. trade balances were beginning to swing to deficits, albeit small ones.[18] Meanwhile, trade as a share of U.S. GDP was also rising. By the 1980s and 1990s, the U.S. was routinely running trade deficits of 3 percent of GDP or more (see figure 1), and trade was a much larger share of the U.S. economy (see figure 2) than it had historically been.

The U.S. trade deficit (or initially, the possibility thereof,) became a controversial issue as early as the Kennedy Round trade negotiations of the 1960s, with the controversy intensifying

[12] For example, Palley explains that Bernanke is ignoring U.S.-based financial activity at the root of financing the U.S. current account deficit, as well as the "barge economics" nature of neoliberal globalization. Palley, "The theory of global imbalances," 2014, at pp. 11-12 and 26-28.
[13] Of course, there are other components to the current account, and those could also be adjusted as part of such proposals.
[14] Benedetto "Who Financed Recent U.S. Trade Deficits?" May 2014.
[15] For example, the issue was raised by famous English economist Joan Robinson, who in 1973 stated that "free trade is just a more subtle form of mercantilism… believed only by those who will gain an advantage from it." (quoted in Lavoie, *Post-Keynesian Economics*, 2014, p. 507). Similarly, in 1977, Robinson described how the British Empire advocated free trade when it had a captive market in its colonies, and then offhandedly asserted that while the United States advocated free trade after World War II, it "has no objection to protection for [its] own industries when they are strongly pressed by Japan." See Robinson, "What are the questions?" December 1977. This author, with the benefit of more historical hindsight than Robinson had in 1977, does not agree that the United States has pursued a vigorously protectionist policy with Japan either before or after 1977. Nor can the enthusiasm of many U.S. policymakers for "free trade" since the 1970s be accurately described as stemming from any U.S. net export advantage, as no such advantage has existed. (This fact does not mean Robinson is wrong about her other historical examples.)
[16] Benedetto "Who Financed Recent U.S. Trade Deficits?" May 2014.
[17] Benedetto "Who Financed Recent U.S. Trade Deficits?" May 2014.
[18] Eckes, *Opening America's Market*, 1995 pp. 197-217.

in the 1980s, and remaining a source of political and economic debate ever since.[19] Since the late 1990s, the major sources of the U.S. trade deficit have widened to include not only Japan and Europe, but also Mexico and especially China.

In sum, a large U.S. trade deficit, accompanied by rising trade overall, has been a constant in the U.S. economy for several decades, in contrast to almost all U.S. economic history before this modern period. U.S. trade deficits have been controversial within the United States, and arguments continue over whether these deficits are a cost to the U.S. economy or a benefit to U.S. consumers and an indicator of relative U.S. economic strength.

Figure 1: U.S. trade and current account balances as a share of GDP, 1960-2015

Source: data from BEA and author's calculations.

Figure 2: U.S. trade (exports plus imports of goods and services) as a percentage of GDP, 1960-2015

Source: data from BEA and author's calculations.

[19] For a discussion of some of the controversies over the U.S. trade deficit in the 1960s and 1970s, see Congressman Don Bonker's *America's Trade Crisis*, 1988, pp. 51-54, Eckes, *Opening America's Market*, 1995 pp. 197-217; Prestowitz *The Betrayal*, 2010, pp. 77-120; and Stein, *Pivotal Decade,* 2010.

U.S. trading partners

But what of U.S. trading partners? How much could the U.S. trade deficit have affected them? This paper will now describe briefly what basic data say has been happening to U.S. trading partners during the same period of time.

As has been shown already, the U.S. trade deficit has been above or around 3 percent of U.S. GDP for many years. At 3 percent of U.S. GDP, the U.S. trade deficit is a much smaller share of the rest of the world's summed GDP. Thus, it is possible that trade surpluses with the United States are not a large factor for most countries, if those trade surpluses with the United States were spread out evenly across the world. However, as shall be shown, for a small number of countries and regions, trade surpluses with the United States are a large share of their annual economic growth.

Table 1 shows the largest U.S. trade deficits (surpluses for the partner country) and surpluses (deficits for the partner country) for the largest partner surpluses and deficits in 2015, as reported by the U.S. Department of Commerce. As can be seen from the table, the largest surpluses dwarf the largest partner deficits. This fact suggests that the U.S. trade deficit is concentrated in a handful of countries/regions: China, the EU, Japan, Korea, and Mexico. Moreover, for the United States, those trade deficits are not cancelled out by large U.S. trade surpluses with other countries. (If they were, then obviously the United States would not be running trade deficits overall.)

Table 1: U.S. trade balance (exports less imports of goods and services) of largest deficit and surplus countries, 2015

Country	**U.S. trade balance (goods and services), billions of dollars**
Countries with a trade surplus with the United States	
China	-334.1
Germany	-77.3
Mexico	-57.9
Japan	-55.4
India	-29.9
Italy	-29.9
South Korea	-18.7
France	-14.7
Taiwan	-10.3
Countries with a trade deficit with the United States	
Hong Kong	31.7
Brazil	25.3
Singapore	17.5
United Kingdom	12.0
Saudi Arabia	6.3
Canada	6.1
All other countries not listed above, together	28.9

Source: BEA data from *U.S. International Trade by Selected Countries and Areas*, table 3, sorted by author.

The basic data in table 1 suggest a follow-up question. How large are these other countries' trade surpluses relative to their own economies?

Table 2 compares the size of those countries' or regions' trade balances (usually surpluses) with the United States to those countries' own GDP, by year. Doing so will allow analysis (below) of how large each country's or region's trade surplus with the United States is, compared to its average annual income (GDP) growth.

Comparing the U.S. trade surplus share of a country's GDP to a country's GDP growth is comparing a flow (the former) to a change in a flow (the latter). The comparison allows a rough idea of the importance of a country's trade surplus with the United States on that country's annual national growth rate. (See Analysis below).[20]

For trade data, table 2 uses both goods and services trade data from the U.S. BEA. For GDP, table 2 uses International Monetary Fund (IMF) data, which is sourced from the countries in question. Thus, table 2 is comparing U.S. trade data with foreign countries' own GDP data. In general, foreign countries' own merchandise trade data is usually close to U.S. trade data.[21] The major exception to this general statement is China, which reports a much lower trade surplus with the United States than the United States reports with it. However, this issue in the Chinese trade data exists not only with the United States, but with many countries. For example, China reports a merchandise trade deficit with Japan every year over 2010-2015, while Japan often reports merchandise trade deficits with China over the same period.[22]

Additionally, a note of interest and a note of caution are warranted for the EU data. First, as the note of interest, Germany and Italy are shown separately in table 2, along with the entire EU. The EU trade surplus with the United States was $102.9 billion in 2015, but the German trade surplus with the United States was $77.3 billion and the Italian trade surplus with the United State was $29.9 billion, suggesting that these two countries represent the lion's share of the EU trade surplus with the United States.[23] Germany and Italy were the first- and fourth-largest economies in the EU in 2015. The United States ran a trade surplus with the EU's second-largest economy (the United Kingdom) and a trade deficit of $14.7 billion with the EU's third-largest economy (France) in 2015.

As the note of caution, the U.S. trade deficit (as reported by the U.S. government) with the EU consists of a large goods trade deficit with the EU, mitigated by a trade surplus with the EU in services. The EU, however, reports a trade surplus in services with the United States.[24] This paper assumes the U.S. services data are correct. However, if the EU services data are correct, then the overall U.S. trade deficit (goods plus services) with the EU is even larger. It is also worth noting that, even according to the U.S. data, the United States runs a trade deficit with Germany and Italy in services (but a surplus with France and the United Kingdom).[25]

[20] This paper is not a modeling exercise designed to weigh multiple possible economic effects against each other if the U.S. trade deficit were to decrease.
[21] Benedetto, "Trade deficits and tax evasion," 2016.
[22] Data from Global Trade Atlas database, accessed August 6, 2015, and author's calculations., and Scott, "Value-added Analysis of Trade," 2013.
[23] See BEA, *U.S. International Trade by Selected Countries and Areas*, September 2015, table 3.
[24] See European Commission Directorate-General for Trade, "Countries and regions United States."
[25] See BEA, U.S. International Trade by Selected Countries and Areas. table 9.

Table 2: Selected large U.S. trading partners' trade surpluses with the United States as a percentage of their own GDP, 1999-2015

	Trade surplus with United States as a percent of GDP								
Year	Canada	China	EU	Germany	Italy	Japan	Korea	Mexico	Saudi Arabia
1999	4.2	6.1	0.3	1.2	1.0	1.3	1.5	3.3	-1.3
2000	6.5	6.9	0.5	1.5	1.3	1.3	2.0	3.1	3.4
2001	6.6	6.2	0.6	1.6	1.3	1.3	2.3	3.5	3.2
2002	5.9	7.0	0.8	1.9	1.2	1.4	1.9	4.5	4.0
2003	5.3	7.4	0.7	1.7	1.0	1.3	1.7	5.1	6.0
2004	6.0	8.3	0.7	1.8	1.1	1.3	2.4	5.5	5.5
2005	6.2	8.8	0.8	1.9	1.2	1.5	1.5	5.2	5.8
2006	4.6	8.4	0.7	1.8	1.2	1.8	1.0	6.1	5.9
2007	3.7	7.3	0.5	1.5	1.0	1.7	0.8	6.6	5.5
2008	3.9	5.8	0.3	1.3	0.9	1.2	0.7	5.3	7.5
2009	0.2	4.3	0.1	0.9	0.7	0.6	0.6	4.6	1.7
2010	0.4	4.3	0.3	1.1	0.7	0.8	0.4	5.6	2.9
2011	0.6	3.7	0.3	1.4	0.9	0.8	0.5	4.9	4.2
2012	0.3	3.5	0.4	1.9	1.1	1.0	0.6	4.6	4.2
2013	0.2	3.1	0.5	2.0	1.1	1.2	0.7	3.8	3.4
2014	0.6	3.0	0.5	2.1	1.3	1.2	1.1	4.0	2.8
2015	-0.4	3.0	0.6	2.3	1.6	1.3	1.4	5.1	-1.0
Average 1999-2015	3.2	5.7	0.5	1.6	1.1	1.2	1.2	4.8	3.6
Average real GDP growth, 1999-2015	2.3	9.4	1.6	1.3	0.3	0.8	4.7	2.4	4.8

Source: data from BEA and IMF, calculations by author

As can be seen from the last two rows of table 2, the U.S. trade deficits with its major trading partners likely have a large impact on the economies of those trading partners. In table 2, over 1999-2015, the average annual contribution of the Canadian, German, Italian, Japanese, and Mexican trade surpluses with the United States to those countries' GDP was more than their own average annual GDP growth. For China, the EU, Korea, and Saudi Arabia, the average annual contribution of their trade surpluses with the United States was a large fraction of their annual growth rate.[26]

This result suggests that for these countries, running a trade surplus with the United States has been an important part of their annual GDP growth for many years. If so, whether one believes the United States has a trade deficit problem or not, these trading partners will also be subject to substantial adjustment should the U.S. trade deficit be reduced.

However, before concluding, this paper now conducts three robustness checks on the table: first, whether the results change if purchasing power parity (PPP) data are used; second, whether energy trade accounts for much of the U.S. trade deficit; and third, whether value-added analysis would change the findings.

[26] By contrast, India, which is not presented in the tables, presented somewhat different results. India's trade surplus with the United States represented an average 1.1 percent of Indian GDP on an exchange rate basis (for 1999-2015), falling to 0.3 percent if PPP was used. These percentages compare to an annual average GDP growth rate of 7.1 percent for India over the same period.

Robustness checks

1. Purchasing power parity

When using another country's GDP data, the question arises of how to compare its data in its own currency to U.S. data in dollars. The IMF data offer both a simple dollar conversion (i.e., the foreign GDP amounts are converted to dollars based on the exchange rate of the time), which are the data used in table 2 above. However, the IMF also offers a conversion based on purchasing power parity (PPP), i.e., by attempting to capture the difference in what each currency can actually purchase in its own country. For countries that are substantially poorer than the United States, PPP conversions may make their GDP look larger than market exchange rate conversions, as their own currencies can buy more in their own countries than a market exchange rate might suggest.

Thus, while this paper takes no position on the relative value of using market exchange rates or PPP rates, table 3 also compares these countries' trade surpluses with the United States to their own GDP, as calculated on a PPP basis. Such a calculation substantially reduces the importance of China's trade surplus with the United States to the Chinese economy from the extremely high level in table 2. However, because the Chinese trade surplus with the United States is so large, the reduction still leaves the Chinese trade surplus with the United States as a high percentage of Chinese GDP- generally over 2 percent and always at least 1.7 percent. PPP data also reduces the size of the U.S. trade surplus on the economies of Mexico and Saudi Arabia, but has much less effect on the other countries in the tables.

Table 3: Selected large U.S. trading partners' trade surpluses with the United States as a percentage of their own GDP on a PPP basis, 1999-2015

	Trade surplus with United States as a percent of PPP GDP								
Year	Canada	China	EU	Germany	Italy	Japan	Korea	Mexico	Saudi Arabia
1999	3.3	2.1	0.3	1.1	0.8	1.8	1.1	1.7	-0.4
2000	5.3	2.3	0.4	1.2	0.9	1.9	1.5	1.7	1.1
2001	5.1	2.0	0.5	1.2	0.9	1.7	1.5	2.1	1.0
2002	4.5	2.3	0.6	1.6	0.9	1.7	1.3	2.7	1.2
2003	4.6	2.5	0.7	1.7	0.9	1.5	1.3	2.8	1.9
2004	5.6	2.9	0.7	1.8	1.1	1.7	1.8	3.0	1.8
2005	6.2	3.1	0.7	1.9	1.1	1.7	1.2	3.1	2.2
2006	5.0	3.1	0.7	1.8	1.1	1.9	0.9	3.7	2.4
2007	4.2	2.9	0.5	1.6	1.0	1.7	0.7	4.1	2.3
2008	4.6	2.6	0.3	1.5	1.0	1.4	0.5	3.4	3.5
2009	0.2	2.0	0.1	1.0	0.7	0.7	0.4	2.5	0.6
2010	0.4	2.1	0.3	1.2	0.7	1.0	0.3	3.3	1.3
2011	0.8	2.0	0.3	1.5	0.9	1.0	0.3	3.0	2.1
2012	0.3	2.0	0.4	1.8	1.1	1.3	0.5	2.7	2.1
2013	0.3	1.8	0.5	2.0	1.1	1.3	0.5	2.3	1.6
2014	0.6	1.8	0.5	2.1	1.3	1.1	0.8	2.4	1.3
2015	-0.4	1.7	0.5	2.0	1.4	1.1	1.0	2.6	-0.4
Average 1999-2015	3.0	2.3	0.5	1.6	1.0	1.5	0.9	2.8	1.4
Average real GDP growth, 1999-2015	2.4	9.4	1.6	1.3	0.3	0.8	4.7	2.4	4.8

Source: data from BEA and IMF, calculations by author.

2. Energy trade

A few of these countries' trade surpluses with the United States are heavily influenced by net exports of energy. Saudi Arabia and Canada, for example, run trade surpluses with the United States only because they run trade surpluses in mineral fuels (as categorized in the Harmonized System code HS 27).[27]

However, for the other large trading partners in the table, trade in HS 27 is a much smaller share of their trade surplus with the United States. For Mexico, its trade surplus with the United States in HS 27 represented only 11.9 percent of its total trade surplus with the United States in 2014, and even fell to a trade deficit with the United States in 2015. However, over 25 percent of Mexico's trade surplus with the United States was from HS 27 from 2009 to 2012. For China, the EU, and Japan, trade balances with the United States in HS 27 are very small shares (less than 3 percent) of their trade surpluses with the United States over 2011-2015. (Additionally, these trade balances are often trade deficits, not surpluses.) Korea, which exports jet fuel (a manufactured product included in HS 27) to the United States, sometimes has its trade surplus in HS 27 rise to almost nine percent of its total trade surplus with the United States because of these exports.[28]

In sum, for the most part, the persistent U.S. trade deficits with most of the largest U.S. trading partners are not energy-related.

3. The impact of value added analysis

Recently, some trade economists are urging that trade relationships between countries be analyzed on a value-added, rather than "gross" (i.e. reported exports and imports) basis. They argue that this analysis would record, for example, Japanese contributions to the products exported by China to the United States as Japanese exports rather than Chinese exports. The most frequently-used value-added data set is the Trade in Value-Added data from the Organization of Economic Cooperation and Development, referred to in this paper as the OECD data.

Some have implied that the countries that run large trade surpluses with the United States on a reported basis are not the countries running large trade surpluses with the United States on a value-added basis.[29] Thus, it may be worthwhile to examine whether this paper's findings above, i.e., that major U.S. trading partners have been running large (reported) trade surpluses (relative to their own economies) for a long time, holds up when examined using the value-added data from the OECD.

However, before doing so, it is also important to note that there are also some concerns about using the OECD value-added data. Creating value-added data involves some estimation to eliminate discrepancies between various countries' trade data.[30] These discrepancies can be large, especially in the case of China, and thus the more general economic reader may not be

[27] This fact is only true recently for Canada; in the mid-2000s, Canada ran an overall trade surplus with the United States net of HS 27 as well.
[28] This paragraph is based on data obtained from Global Trade Atlas.
[29] See Benedetto, "Implications and Interpretations," July 2012 for some of these arguments.
[30] This kind of estimation is true whenever one would try to align different countries' trade data. The concern in this case comes from the apparent proximity of the OECD estimate to the Chinese data, which are often quite different than the data from other countries, as discussed below.

aware of how much the particular value-added data take the Chinese government's data into account.

For example, the Economic Policy Institute (EPI) has pointed out that, with many of its trading partners, China reports different imports and exports than its partners do, and those differences make the Chinese reported trade surplus with the world much smaller than the sum of Chinese trading partners' reported Chinese trade surpluses. While China reported a trade surplus of $157.8 billion with the world in 2011, EPI sums up China's largest trading partners' reported data and finds a Chinese surplus with the world of $526.3 billion in the same year. EPI then notes that the OECD data state that China ran a merchandise trade surplus with the world of $187.7 billion in 2011, obviously much closer to the Chinese government's estimate than to the rest of the world's estimate. My paper is not endorsing one estimate or another, but notes that there are these widely variant descriptions of the Chinese trade relationship with the world that come from whether one uses data from the Chinese government to some substantial extent (OECD) or not at all (EPI).[31] [32]

However, even if one leaves these concerns aside, this paper now considers how the OECD's value-added data might change this paper's findings above. The most recent value-added data available from the OECD are from 2011. From what these data show, the U.S. trade deficit with China on a value-added basis would be somewhat smaller than on a reported basis, but would still be large. For example, in 2011, according to the OECD value-added data, China had total value-added exports to the world of $2.0 trillion, of which $1.3 trillion (68 percent) was Chinese value-added. However, an additional $0.3 trillion of those Chinese total value-added exports were value-added in Canada, the EU, Japan, Korea, and Mexico. Overall, 81 percent of Chinese value-added exports were value-added from China or those countries.[33]

While value-added data can reallocate the reported U.S. trade deficit among its trading partners, value-added data cannot change the total value of the U.S. trade deficit in theory.[34] (In practice, the OECD estimate that seems to lend a lot of weight to the Chinese data would reduce the U.S. trade deficit. However, this estimated reduction is entirely due to the use of Chinese government data, and not because of the use of value-added analysis.) Thus, even to the extent that China's trade surplus with the United States is somewhat lower in the OECD data, it appears the main beneficiaries (in terms of exporting value to China that the United States imports in Chinese products) would mostly be the other countries already discussed.
Additionally, it is important to remember that China has been moving up the value chain both before 2011 and since, and could be adding more value now than it was before. However,

[31] See Scott, "Value-added Analysis of Trade," 2013. Some economists have also raised concerns about how other parts of Chinese GDP are measured, and argued that Chinese GDP overall is overestimated. See, for example, Rawski, "What is happening to China's GDP statistics?" 2001.
[32] Furthermore, value-added data are based on an assumption that labor costs in all countries reflect market conditions. See McMillion, "China Trade Apologists" October 2011; Benedetto, "Implications and Interpretations," July 2012; Scott, "Value-added Analysis of Trade," 2013; and Mandel "Introduction," 2015. As Benedetto (2012) explains, foreign government subsidies and/or foreign labor conditions can affect how much measured "value-added" is in an export. Mandel (2015) also notes this point.
[33] Data for this analysis from OECD Trade in Value-Added Database. Calculations are the author's. It should also be noted that in the OECD data, U.S. exports of value-added are also a little lower than U.S. reported exports, so U.S. exports would also be reduced if the OECD data were used. (Thus, lower Chinese value-added in its exports to the United States, which would make the U.S. value-added trade deficit with China lower than the U.S. reported trade deficit with China, would be partially offset by lower U.S. value-added in its exports to China, which would somewhat increase the magnitude of the U.S. value-added trade deficit with China.)
[34] See Benedetto, "Implications and Interpretations," July 2012.

even if China were only adding what the value-added data report for 2011, the U.S. trade deficit would remain large, and a large percentage of Chinese GDP.

Thus, it is highly likely that (1) most (at least two-thirds)[35] of Chinese exports to the United States consists of Chinese value-added and (2) even if as large a share of Chinese exports to the United States consists of value-added in other countries as the OECD estimates, those other countries are primarily the other ones that run persistent large trade surpluses with the United States.

In other words, value-added analysis using the OECD data would almost leave the trade surplus with the United States as a large contributor to the GDP of China, the EU, Japan, and Korea, albeit with somewhat less emphasis on China and more on the other large U.S. trading partners.

Analysis

The large share of average annual GDP growth accounted for by several countries' trade surpluses with the United States suggests that reducing the U.S. trade deficit would require not only a large adjustment for the United States, but also for these nations. The basic story remains the same whether nominal exchange rates or PPP rates are used. Use of value-added data may change the exact amounts for each trading partner, but not the basic story. If many large U.S. trading partners were to lose their typical trade surplus with the United States, they would need to make up a substantial share of their typical GDP growth somewhere else.

Modern economists might argue that it is the United States that has benefitted from the imports from these nations, and that these nations would be better off without such an emphasis on net exports to the United States. Even so, modern economists could find that these countries would also require large adjustments to their own economies if they were to lose their persistent trade surpluses with the United States.

On the other hand, economic analysis rooted in an older Keynesian tradition might lead one to conclude that these nations have grown accustomed to a boost to their own economies from large net exports to the United States. Such an analysis might argue that these nations need to reduce their surpluses in order to maintain a harmonious global economy.[36]

Whatever one's view of trade balances and their costs and benefits, it is clear that the U.S. trade deficit is not an issue of relevance only to the United States. For many years, the United States' largest trading partners have been receiving a large portion of their annual GDP from their repeated trade surpluses with the United States.

[35] See also OECD-WTO *Trade in Value-Added Data*, October 2015. If one were to use the two-thirds estimate noted on U.S. data on goods trade with China, it would reduce the Chinese trade surplus with the United States from from $367 billion in 2015 to $207 billion, still quite large. This calculation assumes that one-third of 2015 U.S. imports from China reflect other nations' value-added, and no 2015 U.S. exports to China reflect Chinese value-added. Whether or not these assumptions are reasonable, I use them to show that even under the most generous assumptions possible, the Chinese trade surplus with the United States remains large.

[36] Keynes himself advocated such an adjustment by surplus countries as part of the Bretton Woods system. See Lavoie, *Post-Keynesian Economics*, 2014, p. 500.

Conclusions

This paper has presented relatively basic economic data to show that the U.S. trade deficit is not only a large-scale economic reality to the U.S. economy, but also (as its mirror, a trade surplus) to the largest U.S. trading partners. The paper has done so by comparing U.S. trade balance data with the GDP data of other nations. It has taken into account potential wrinkles such as purchasing power parity and value-added analysis, and found that while such analysis could change the precise size or distribution of other countries' trade surpluses with the United States, it would not change the fundamental conclusion.

These basic data strongly suggest that reducing the U.S. trade deficit would likely require or force substantial adjustments by U.S. trading partners. To the extent policymakers in these countries use traditional Keynesian analysis, the loss of their trade surplus with the United States would mean a large adjustment to their national economies. Modern economic analysis, which might conclude that other countries' large trade surpluses with the United States are not a long-run benefit to those countries, might also conclude that there could be large adjustments for those countries as well, albeit perhaps more to their own benefit. Either way, it is highly possible that the scale of these adjustments would cause major geopolitical ripples. It is fair to say the U.S. trade deficit is not only a U.S. issue.

Bibliography

Baker, Dean. "'Secular Stagnation' Is Progress, Now Can We Talk About the Trade Deficit?" *Beat the Press* blog, Center for Economic and Policy Research, November 2014. http://cepr.net/blogs/beat-the-press/qsecular-stagnationq-is-progress-now-can-we-talk-about-the-trade-deficit

Benedetto, John B. "Implications and Interpretations of Value-Added Trade Balances." *Journal of International Commerce and Economics,* July 2012.

Benedetto, John B. "Who Financed Recent U.S. Trade Deficits?" *Journal of International Commerce and Economics*, May 2014.

Benedetto, John B. "Trade deficits and tax evasion," *Journal of International Commerce and Economics*, forthcoming.

Bernanke, Ben S. "Remarks," The Sandridge Lecture, Virginia Association of Economists, Richmond, Virginia, March 10, 2005

Bernstein, Jared. "What we're arguing about when we're arguing about trade deals." *Jared Bernstein On the Economy Blog,* May 18, 2015. http://jaredbernsteinblog.com/what-were-arguing-about-when-were-arguing-about-trade-deals/

Bivens, Josh. "Yes, Trade Deficits Do Indeed Matter for Jobs," *Working Economics Blog*, Economic Policy Institute, May 28, 2015. http://www.epi.org/blog/yes-trade-deficits-do-indeed-matter-for-jobs/

Bonker, Don. *America's Trade Crisis*. Houghton-Mifflin Company, 1988.

Eckes, Alfred E. *Opening America's Market*. The University of North Carolina Press, 1995.

European Commission Directorate-General for Trade, "Countries and regions United States." http://ec.europa.eu/trade/policy/countries-and-regions/countries/united-states/ accessed February 4, 2016.

Global Trade Atlas database, accessed August 6, 2015.

International Monetary Fund (IMF), *World Economic Outlook Database*, April 2016.

Ip, Greg. “The Case for Free Trade Is Weaker Than You Think,” April 11, 2016, http://blogs.wsj.com/economics/2016/04/11/the-case-for-free-trade-is-weaker-than-you-think/ accessed May 5, 2016.

Irwin, Douglas A. *Against the Tide*. Princeton, New Jersey: Princeton University Press, 1996.

Keen, Steve. *Debunking Economics*. New York: Zed Books, 2001.

Lavoie, Marc. *Post-Keynesian Economics: New Foundations.* Northampton: Edward Elgar Publishing, 2014.

Mandel, Michael. “Introduction.” In *Measuring Globalization: Better Trade Statistics for Better Policy.* Vol. 2. *Factoryless Manufacturing, Global Supply Chains, and Trade in*

Intangibles and Data, edited by Susan N. Houseman and Michael Mandel. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2015.

McMillion, Charles. “China Trade Apologists Know a ‘Reality’ That is NOT True.’ *Manufacturing and Technology News.* Volume 18, No. 16. October 21, 2011.

Milberg, William. “Say’s law in the open economy: Keynes’ rejection of the theory of comparative advantage.” in *Keynes, Uncertainty, and the Global Economy* ed. Sheila Dow and John Hilliard, Edward Elgar Publishing, Northampton, MA (2002).

Organization for Economic Cooperation and Development (OECD), *Trade in Value-Added Database*, http://stats.oecd.org/Index.aspx?DataSetCode=TIVA2015_C1 downloaded March 12, 2016.

Organization for Economic Cooperation and Development (OECD)- World Trade Organization (WTO). *Trade in Value-Added: China*, October 2015.

Palley, Thomas. “The theory of global imbalances: mainstream economics vs. structural Keynesianism.” *Institut für Makroökonomie und Konjunkturforschung Macroeconomic Policy Institute*, Working Paper, September 2014.

Perry, Mark J. “Imports and trade deficits do not ‘hinder’ economic growth and are not a ‘setback’ for the US economy,” AEIdeas blog, May 6, 2015. http://www.aei.org/publication/imports-and-trade-deficits-do-not-hinder-economic-growth-and-are-not-a-setback-for-the-us-economy/

Prestowitz, Clyde. *The Betrayal of American Prosperity*. New York: Free Press. 2010.

Rawski, Thomas. “What is happening to China’s GDP statistics?” *China Economic Review* 12 (2001) 347-354.

Robinson, Joan. “What Are The Questions?” *Journal of the Economic Literature*, Vol. 15, No. 4 (December 1977), pp. 1318-1339.

Schumacher, Reinhard. “Deconstructing the Theory of Comparative Advantage.” *World Economic Review* 2:83-105, 2013.

Scissors, Derek. “The trade deficit does not cost us jobs,” AEIdeas blog, March 16, 2015. http://www.aei.org/publication/the-trade-deficit-does-not-cost-us-jobs/

Scott, Robert E. “Value-added Analysis of Trade with China Could Weaken Fair Trade Enforcement and Increase Job Loss.” Economic Policy Institute Briefing Paper no. 361, June 27, 2013.

Smith, Noah. “How ‘Keynes’ Became a Dirty Word,” Bloomberg.com, September 11, 2014, https://www.bloomberg.com/view/articles/2014-09-11/how-keynes-became-a-dirty-word

Smith, Yves. *Econned*. New York: Palgrave and Macmillan, 2010.

Stein, Herbert. "Balance of Payments." The Concise Encyclopedia of Economics. 2008. Library of Economics and Liberty. Retrieved February 3, 2016 from the World Wide Web: http://www.econlib.org/library/Enc/BalanceofPayments.html

Stein, Judith. *Pivotal Decade: How the United States Traded Factories for Finance in the Seventies*. New Haven and London: Yale University Press, 2010

U.S. Department of Commerce (USDOC). Bureau of Economic Analysis (BEA). U.S. Gross Domestic Product, August 26, 2016

U.S. Department of Commerce (USDOC). Bureau of Economic Analysis (BEA). U.S. International Trade by Selected Countries and Areas, December 4, 2015.

U.S. Department of Commerce (USDOC). Bureau of Economic Analysis (BEA). U.S. International Trade in Goods and Services, June 3, 2016.

U.S. Department of Commerce (USDOC). Bureau of Economic Analysis (BEA). U.S. International Transactions, Table 1.1, June 16, 2016.

Author contact: jbnere@verizon.net

SUGGESTED CITATION:
John B. Benedetto, "A note on recent data on U.S. trading partners' trade surpluses with the United States", *real-world economics review*, issue no. 77, 10 December 2016, pp. 103-117,
http://www.paecon.net/PAEReview/issue77/Benedetto77.pdf

real-world economics review, issue no. 77
subscribe for free

A CasP model of the stock market

Shimshon Bichler and Jonathan Nitzan[1] [Israel and York University, Canada]

Abstract
Most explanations of stock market booms and busts are based on contrasting the underlying 'fundamental' logic of the economy with the exogenous, non-economic factors that presumably distort it. Our paper offers a radically different model, examining the stock market not from the mechanical viewpoint of a distorted economy, but from the dialectical perspective of capitalized power. The model demonstrates that (1) the valuation of equities represents capitalized power; (2) capitalized power is dialectically intertwined with systemic fear; and (3) systemic fear and capitalized power are mediated through strategic sabotage. This triangular model, we posit, can offer a basis for examining the asymptotes, or limits, of capitalized power and the ways in which these asymptotes relate to the historical and ongoing transformation of the capitalist mode of power.

1. Introduction

The purpose of this paper is to outline a capital-as-power, or CasP, model of the stock market. There are two reasons why such a model is needed: first, the stock market has become the main compass of the capitalist mode of power; and, second, so far, we have not developed a CasP theory to describe it.[2]

Surprising as it may sound, all long-term modeling of the stock market derives from a single meta-dogma that we have previously dubbed the 'mismatch thesis' (Bichler and Nitzan 2009, 2015a). The basic premise of this dogma is the general bifurcation between economics and politics (a shorthand for all non-economic realms of society) and the further division, within economics, between the so-called 'real' and 'nominal' spheres. Finance in this dogma is a symbolic nominal mirror that reflects the underlying real economy, but that reflection – and this is the key point here – is imperfect. Financial magnitudes tend to mismatch reality, and the purpose of the model is to explain this mismatch and predict its consequences.[3]

Our CasP model begins not by negating these conventional findings and predictions, but by giving them a totally different interpretation. The model suggests that underneath the

[1] This paper was presented at the fourth CasP conference, 'Capital as Power: Broadening the Vista' (York University, September 28-30, 2016). Shimshon Bichler teaches political economy at colleges and universities in Israel. Jonathan Nitzan teaches political economy at York University in Canada. All of their publications are available for free on *The Bichler & Nitzan Archives*. Work on this paper was partly supported by the SSHRC.

[2] The theory of capital as power posits that capitalism is best understood not as a mode of consumption and production, but as a mode of power, and that capital is not machines or labour time, but the central power institution of capitalism. For readers new to the subject, Bichler and Nitzan (2012b) provides a short summary of CasP, Nitzan and Bichler (2009a) offers an extended articulation and Bichler and Nitzan (2015c) brings together a collection of recent articles. The past, present and future of the CasP project, including an extensive bibliography, are outlined in Bichler and Nitzan (2015b).

[3] It is of course true that many practical investment models – such as those based on momentum, quantitative analysis and index tracking, among other rituals – do not rest directly on the valuation of economic fundamentals. But all of them assume that, in the final analysis, capitalization mirrors and must eventually converge to these fundamentals. For bestselling manuals on valuation, see Damodaran (2011, 2012), McKinsey & Company *et al.* (2005) and McKinsey & Company *et al.* (2011).

economic veneer of the mismatch thesis lies a power process, and that it is this power process – and not economic productivity and utility – that drives the stock market. This alternative interpretation is important for three reasons: first, it gives rise to questions that conventional theories are unable to ask; second, it leads to findings that contradict some of the underlying assumptions of both mainstream and heterodox political economy; and, third, it might open the door for a better understanding of the capitalist mode of power and how it might be resisted and transformed.

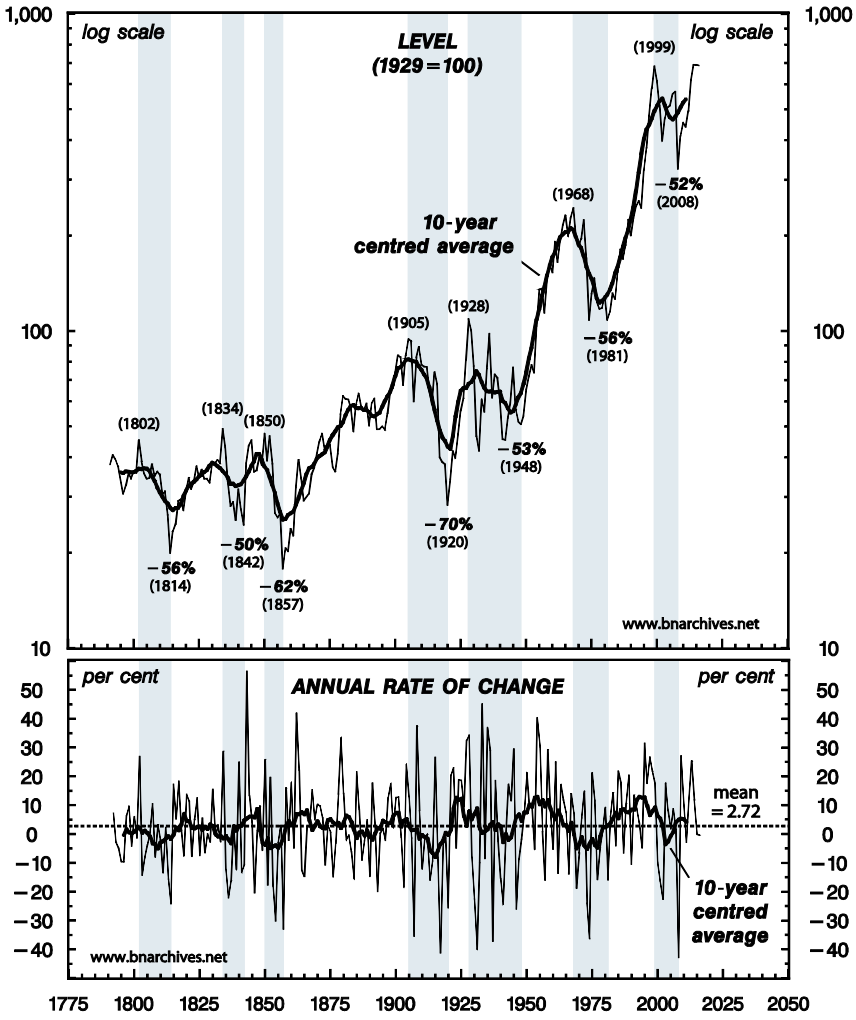
The paper consists of eight sections. The substantive discussion begins in Section 2 with a bird's eye view of stock-market booms and busts over the past two centuries. This section identifies some of the market's quantitative patterns along with the qualitative power transformations that underlie them. Section 3 explains the mainstream mismatch thesis, while Section 4 describes the valuation model of John Hussman, President of the *Hussman Investment Trust*, which, as far as we know, offers the best consistent predictions of long-term stock market returns. The remainder of the paper outlines our own model, illustrated by the enclosed Penrose triangle. Section 5 shows that one can reproduce Hussman's results by looking not at the utilitarian economics of production and consumption, but directly at *capitalized power*. Section 6 explores how capitalized power is dialectically intertwined with what we call *systemic fear*. Section 7 suggests that the driving force behind both capitalized power and systemic fear is what Thorstein Veblen called *strategic sabotage* and speculates on how economic policy has been integrated into the CasP-driven stock market. Section 8 concludes with a brief summary and some thoughts about the future.

Before turning to our argument, we should note that, although the principles we deal with here are general, our empirical analysis is restricted to the United States. Given this limitation, our article should be taken as explorative and tentative rather than exhaustive and definitive.

2. Major bear markets

Begin with Figure 1, which shows annual U.S. stock prices since the late eighteenth century. The top panel displays levels, while the bottom panel shows rates of change. Prices are expressed in 'constant dollars', which means that, for every year, the nominal stock price index is divided by the CPI (Consumer Price Index).

Figure 1: U.S. Stock Prices in Constant Dollars, 1791-2016



NOTE: Shaded areas indicate major bear markets (MBMs) as defined in the text and in Table 1. Negative numbers in the top panel indicate the decline of the CPI-adjusted market price from the MBM's peak to trough (trough year in parentheses). The U.S. stock price index splices the following four sub-series: a combination of bank, insurance and railroad stock series weighed by Global Financial Data (1820-1870); the Cowles/Standard and Poor's Composite (1871-1925); the 90-stock Composite (1926-1956); and the S&P 500 (1957-present). The constant-dollar series is computed by dividing the stock price index by the Consumer Price Index (CPI). Data are rebased with 1929=100.0. The last data point is 2016 for the underlying series and 2011 for the ten-year centred average.

SOURCE: Stock prices are from Global Financial Data (GFD) till 1900 (series codes: _SPXD) and from Standard and Poor's through Global Insight (GI) from 1901 onward (series codes: JS&PNS). The CPI is from GFD till 1947 (series code CPUSA) and from the IMF's International Financial Statistics through GI from 1948 onward (series code: L64@C111).

2.1 Purchasing power or differential power?

Economists go through this transformation in order to display stock prices in so-called 'real terms'. And what they mean by 'real terms' here is purchasing power.[4] According to this view,

[4] Much of economics is conceived, theorized, measured and written in 'real terms'. We enclose this concept in inverted commas because, in our view, it is deeply problematic both philosophically and empirically (see for example Nitzan 1989; Nitzan and Bichler 2009a: Ch. 8).

each reading on the thin line in the top panel of the chart shows the purchasing power, denominated in universal 'utils', of the owners of the largest U.S.-listed firms.[5] For example, if the thin line rises by 10 per cent, that means that stock owners like Buffett and Soros can buy 10 per cent more groceries (measured in utils), or purchase 10 per cent more fuel for their vehicles (again, weighted in utils). Conversely, if this line drops by 30 per cent the implication is that the Buffetts and Soroses of the world can afford 30 per cent less clothing or can spend only 30 per cent less on childcare (both counted in utils).

Surprising as it may seem, this is the dominant view. In a hyper-liberal universe, goes the argument, every individual, whether a billionaire owner or a propertyless beggar – and by extension, every association of individuals, be it a corporation, a government, or an NGO – is ultimately driven by an innate desire to maximize hedonic pleasure, and the compulsive conversion of all nominal measures into 'real terms' is a way of guiding and gauging this utilitarian obsession.

But there is another, perhaps more meaningful way to think about this computation. When we divide stock prices by the CPI, we are computing a ratio between the prices of two baskets: a basket of listed corporate stocks and a basket of consumer goods and services. And if we think of relative commodity prices as reflecting the relative power of their owners, then what we measure here is not only purchasing power, but also differential power: the power of those who own stocks relative to the power of those who own consumer goods and services. So we have two measures in one: purchasing power and differential power, and in what follows we focus solely on the latter.

2.2 The stylized facts

So what does Figure 1 show us? First, it shows that, since the late eighteenth century, the differential power of stock owners relative to owners of consumer goods and services has risen exponentially (notice that the top panel uses a log scale, so equal vertical distances represent multiples of 10). The mean geometric growth rate for the entire period is 1.28 per cent.[6] If this number seems small, note that over the past 215 years it has compounded to a nearly 18-fold rise in the relative power of stock owners. The bottom panel of the chart shows the annual rate of change. The thin series in this panel displays, for each year, the growth of U.S. stock prices in constant dollars relative to the previous year (with the mean arithmetic growth rate of 2.72 per cent indicated by the horizontal dashed line).

Each panel also show a thick series. In each case, the series displays the ten-year centred average of the respective thin series. Every observation in the thick series is the mean value of the respective thin series (level or rate of change) computed over a ten-year window. Since the moving average is centred, the window consists of the five years preceding the

[5] The term 'util' was invented by Irving Fisher (1892). Since this unit is totally fictitious and impossible to measure, contemporary mainstream economists pretend it to be theoretically unnecessary. In practice, though, they conveniently forget this pretention and use the util in pretty much everything they do. The reason for this organized, church-like hypocrisy is simple: the only way for economists to compute real economic aggregates (like real GDP and the real capital stock) or real economic ratios (such as those comparing the purchasing power parity or productivity of different economic agents) is to assume a universal unit of measurement; and in the hedonic world of economics, the only universal unit of measurement is the fictitious util (Nitzan 1992: Ch. 5; Nitzan and Bichler 2009a: Chs. 5 and 8).

[6] The geometric growth rate is derived by taking the ratio between the last and first observation, raising this ratio to the power of $1/n$ (with n = number of years in the series less one), subtracting one and multiplying the result by 100.

observation and the five years following it. For example, the ten-year centred average for 1990 is the mean value over the 1986-1995 period.

The thick series in the top panel is useful in describing what we call 'major bear markets', or MBMs.[7] In a market whose price trends upwards exponentially, the most dramatic event is a protracted crash, or MBM. We define an MBM as a period during which the following two conditions apply: (1) the ten-year centred average of the market price, expressed in constant dollars, is declining; and (2) every peak of the annual series is followed by a lower peak (note that the peak/trough of an MBM can slightly precede/trail the inflection point of the ten-year centred average). Based on this dual definition, we can count seven MBMs (shaded in grey): three in the early half of the nineteenth century, three in the twentieth and one in the twenty-first. The top panel of Figure 1 marks the peak and trough years of every MBM, along with the total stock price drop in CPI-adjusted terms. This information is summarized in Table 1.

Table 1: Major U.S. Bear Markets*
(constant-dollar calculations)

Period (Peak-Through)	*Decline (%)*
1802–1814	–56%
1834–1842	–50%
1850–1857	–62%
1905–1920	–70%
1928–1948	–53%
1968–1981	–56%
1999–2008	–52%

* A major bear market (MBM) is defined as a multiyear period during which: (1) the ten-year centred average of stock prices, expressed in constant dollars, trends downward; and (2) each successive sub-peak of the underlying price series, expressed in constant dollars, is lower than the previous one. Note that the peak/trough of an MBM can slightly precede/trail the inflection point of the ten-year centred average.

2.3 The quantities and qualities of power

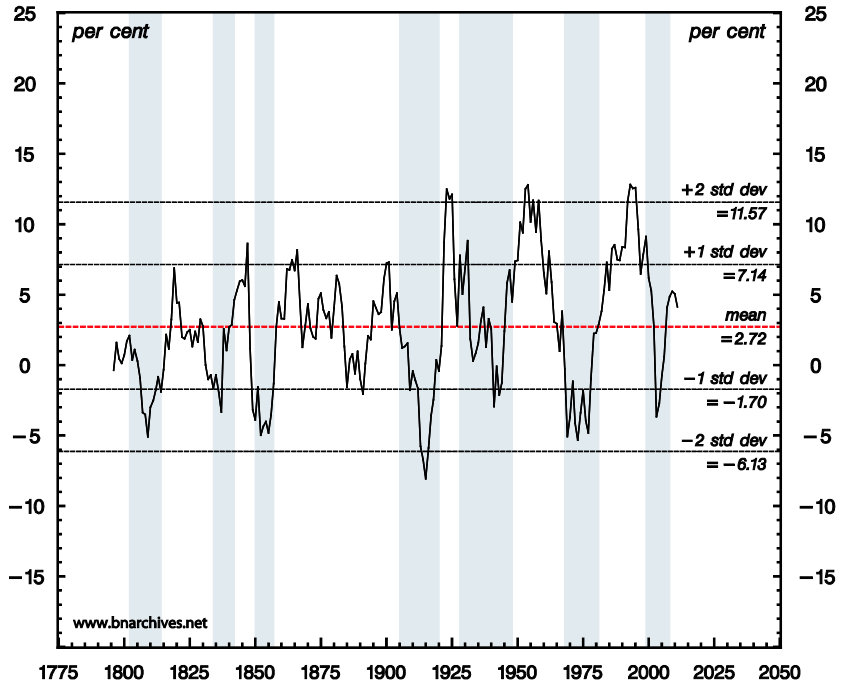
Now, when the market is booming, the popular media loves to insist that 'this time is different'. The talking heads claim that something fundamental has changed, and that the good times can roll on forever. This point was infamously made in October 1929 by Irving Fisher, just before he lost his fortune, equivalent to $100 million in today's prices, to the Great Crash (Anonymous 1929), and then again in 1999 by a couple of far-sighted analysts, who predicted that the stock market would triple in five years, only to see it halved (Glassman and Hassett 1999).

[7] For the genesis, earlier versions and prior analyses of this concept, see Bichler and Nitzan (2008), Kliman, Bichler and Nitzan (2011), and Bichler and Nitzan (2012a).

The pundits like to ridicule these 'new-order' forecasts with counter-titles such as *Irrational Exuberance* (Shiller 2000) and tongue-in-cheek phrases like *This Time is Different* (Reinhart and Rogoff 2009). The market, they say, has its own natural, mean-reverting pattern, and there is nothing anyone can do to change it. And on the face of it, they seem to have a point.

Figure 2 magnifies the lower panel of Figure 1, ignoring the year-on-year variations and showing only the ten-year centred average. This larger and cleaner exposition helps us discern three general patterns: (1) the growth rate of stock prices is not even, but mean-reverting; (2) it is not random, but cyclical; and (3) its cycle is not haphazard, but instead displays a fairly stable duration (though its amplitude in the twentieth century is twice that of the nineteenth's). On the whole, then, the process certainly resembles a naturally reoccurring phenomenon.

Figure 2: Annual Rate of Change of U.S. Stock Prices in Constant Dollars
(ten-year centred average)



NOTE: Shaded areas indicate major bear markets (MBMs) as defined in the text and in Table 1. The U.S. stock price index splices the following four sub-series: a combination of bank, insurance and railroad stock series weighed by Global Financial Data (1820-1870); the Cowles/Standard and Poor's Composite (1871-1925); the 90-stock Composite (1926-1956); and the S&P 500 (1957-present). The constant-dollar series is computed by dividing the stock price index by the Consumer Price Index (CPI). The last data point for the ten-year centred average is 2011.

SOURCE: Stock prices are from Global Financial Data (GFD) till 1900 (series codes: _SPXD) and from Standard and Poor's through Global Insight (GI) from 1901 onward (series codes: JS&PNS). The CPI is from GFD till 1947 (series code CPUSA) and from the IMF's International Financial Statistics through GI from 1948 onward (series code: L64@C111).

But is it? Notice that, during the past century or so, every MBM was followed by a major *creordering* of capitalized power and a significant rewriting of the capitalist *nomos*.[8] Thus, the MBM of 1905-1920 was followed by the rise of *corporate capitalism*; the MBM of 1928-1948 was followed by the rise of the *Keynesian welfare-warfare state*; and the MBM of 1968-1981 was followed by the rise of *global neoliberalism*. In this regard, the first MBM of the twenty-first century, from 1999 to 2008, seems incomplete: although the capitalist system is already rattling, it might take another MBM to bring about a significant creordering similar to the previous three.

Now the key thing about these creorderings is that they are *qualitatively different* from each other. So in the end, we can say that 'this time – like every other time – is *both* similar and different': the similar quantities of power are driven by different qualities of power.

3. The mismatch thesis

How should we make sense of these long-term patterns? As noted, the common and perhaps only theory on offer is the mismatch thesis. According to this thesis, capitalism – like any other economic system – is a mode of production and consumption that can be described and analysed in 'real terms' (i.e. in utils). When economists say that the economy grew by 5 per cent, what they mean – whether they are aware of it or not – is that there were 5 per cent more utils produced this year than last. This is the so-called real sphere of the economy.

But the capitalist economy has another, financial sphere, and that sphere is denominated not in real utils, but in nominal dollars and cents. The stock market is part of this financial sphere, and it plays a crucial role: it acts as a giant mirror that reflects what happens – or, to be more precise, what *will* happen – in the real sphere.

3.1 House of mirrors

Picture 1 illustrates how this mirror operates. According to the mismatch thesis, which can be traced back to Irving Fisher (1896, 1907), capitalists own machines, raw materials and knowledge that together make up the 'real capital stock' (top left quadrant). This real capital stock will produce, in the future, real income services (top right) that the capitalists will in turn sell for nominal future profit (bottom right). And the expectations of this future profit are capitalized by their owners, here and now, into present nominal market value (bottom left).[9] Now if these transmutations are accurate, we have a match; if they are inaccurate, we have a mismatch.[10]

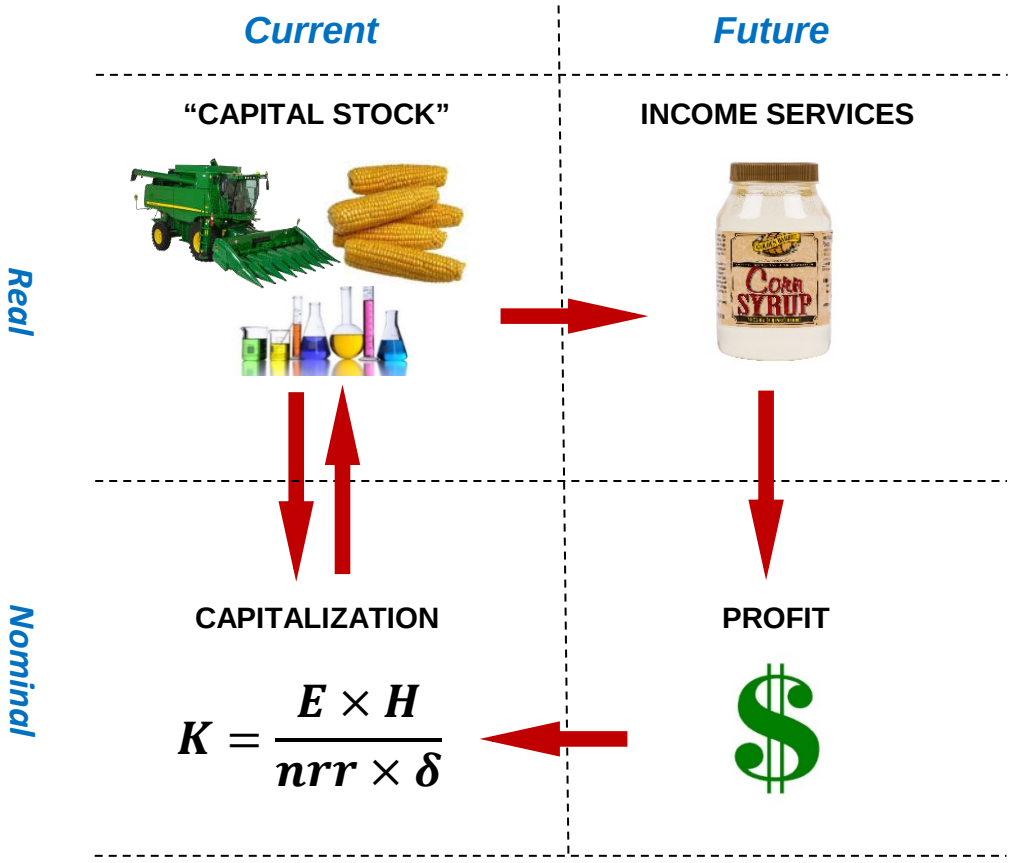
---

[8] The verb-noun 'creorder' fuses the dynamic and static aspects of creating order (Nitzan and Bichler 2009a, especially Ch. 14). The word 'nomos' was used by the ancient Greeks to denote the broader social–legal–historical institutions of society (Castoriadis 1984, 1991). The capitalist nomos is explored in Nitzan and Bichler (2009a: Ch. 9).

[9] According to Irving Fisher, 'The statement that "capital produces income" is true only in the physical sense; it is not true in the value sense. That is to say, *capital-value does not produce income-value.* On the contrary, income-value produces capital value. . . . [W]hen capital and income are measured in *value*, their causal connection is the reverse of that which holds true when they are measured in *quantity.* The orchard produces the apples; but the value of the apples produces the value of the orchard. . . . We see, then, that present capital *wealth* produces future income-*services*, but future income-*value* produces present capital-*value*' (1907: 13-14, original emphases). For a critical assessment of Fisher's framework, see Nitzan and Bichler (2009a: 170-172).

[10] Although we do not pursue the following argument in our paper here, we should note that capitalization can neither match nor mismatch the so-called real capital stock, and for the simplest of

Picture 1: Fisher's House of Mirrors



3.2 Capitalization, fundamentals and mean reversion

Let us now narrow our focus and examine the final stage of this process – namely, the process of capitalization – using the following notations in conjunction with Equations 1-3:

E	future profit ($)
GVA	gross value added ($)
H	hype (decimal)
K	market capitalization ($)
m	coeff. reflecting the future mean profit share E/GVA and future mean GVA growth
nrr	normal rate of return (decimal)
δ	risk (decimal)

reasons: this stock does not – and *cannot* – have a definite quantity to start with. As it stands, the capital stock comprises various entities, such as machines, structures, raw materials and (some say) knowledge, yet these entities have *no common unit* with which they can be aggregated into a single magnitude (recall that the universal util is a theoretical and empirical fiction – see footnote 5). This impossibility was first pointed out by Veblen and Wicksell at the turn of the twentieth century, demonstrated by the Cambridge Capital Controversies of the 1950s and 1960s, and reluctantly confirmed by the leading lights of mainstream economics shortly thereafter. However, since accepting this conclusion would have pulled the rug from under the entire edifice of economics, economists continue to insist that capital does have a 'real' quantity and that their discipline, although based on a logical-empirical impossibility, is still a science (for critical summaries, see Harcourt 1969, 1972; Nitzan and Bichler 2009a: Chs. 5 and 8; Hunt and Lautzenheiser 2011: Ch. 16).

Equation 1 conceptualizes the way in which investors capitalize future profit:

1. $K = \frac{E \times H}{nrr \times \delta}$

In this equation the numerator represents expected future profit, while the denominator is the discount rate. The E in the numerator is the actual future profit that *will* be earned. Since this future profit cannot be known here and now, capitalists have to guestimate it; and, usually, their projections are either too optimistic or too pessimistic. This optimism/pessimism is captured by the hype coefficient H. When capitalists happen to be right on target, $H = 1$. This situation, though, is exceptional. Normally, capitalists are either overhyped, so $H > 1$, or underhyped, so $H < 1$. The net effect of hype is to increase or decrease the profit that owners discount *here and now* relative to the profit they (or subsequent owners) will actually earn *in the future*.

The denominator, which represents the discount rate, is also made up of two components. The first component is the normal rate of return nrr, which capitalists believe they are entitled to when investing in so-called riskless assets (such as U.S. government bonds held to maturity). The second component – the scaling factor δ – accounts for the additional return that capitalists demand in order to compensate for the relative riskiness of the particular asset in question (in this case, the risk of equities relative to U.S. government bonds).[11] For example, if the normal rate of return nrr is 0.05, or 5 per cent, and stocks are deemed twice as risky as government bonds, δ will be 2 and the discount rate will be 0.1, or 10 per cent.

The problem with Equation 1 is that profits oscillate violently, so, for the purpose of prediction, it is convenient to take a simplifying shortcut. Instead of estimating future profit directly, capitalists and analysts often start with the overall dollar value of production, or gross value added GVA, and then introduce two guestimates. Looking all the way into the deep future, they project that (1) GVA will grow at a certain average rate, say 5 per cent; and (2) that the share of profit in GVA will oscillate around a certain average, say 6 per cent. Packing these two guestimates into a coefficient m, they can then derive market capitalization by looking at GVA and m instead of E, with the hype coefficient H denoting their excessive optimism/pessimism regarding the magnitude of m. This new computation is shown in Equation 2:

2. $K = \frac{GVA \times m \times H}{nrr \times \delta}$

Equation 3 rearranges the terms of Equation 2. It shows that market capitalization is the product of two components: (1) GVA, which is relatively unambiguous, regularly measured and broadly agreed upon; and (2) the product/ratio of the remaining four components, which comprises a hodgepodge of guesses, hard-to-pin-down conventions and shifting beliefs.

3. $K = GVA \times \frac{m \times H}{nrr \times \delta}$

This decomposition leads us to the gist of the mismatch thesis. If the future were known, the four-element product/ratio on the right-hand side of Equation 3 would probably sum up to a

[11] The risk coefficient reflects the relative confidence capitalists have in their earnings predictions – in this case, the prediction of corporate profit as opposed to interest payments on government debt.

constant: m is pretty much a fixed number (by definition), nrr would be a fixed number equal to the trend growth of GVA and, since there is no uncertainty, both hype H and the risk coefficient δ would be 1.

As we noted, though, the future is not only unknown, but unknowable; and when capitalists and analysts deal with the unknowable, the regrettable result, or so we are told, is uncertainty, irrationality and distortions. All in all, then, the second component on the right-hand side of Equation 3 is sure to oscillate. But this oscillation, the theory's advocates assure us, is part and parcel of the mismatch thesis: while uncertainty, irrationality and distortions abound, they are *not unbounded*. Although the movement of the second element on the right-hand side of Equation 3 may have a large amplitude and a long duration, it is ultimately self-correcting – or, in the more formal language of statistics, *mean-reverting* – and this mean-reversion is the theoretical basis of all long-term valuation models.

4. Hussman's mismatch model

So let us now see how these principles pan out in practice. Our focus in this section is on the model of fund manager John Hussman, which, as far as we know, offers the best prediction of long-term future returns.[12]

4.1 Hussman's mismatch index (HMI)

The model is based on Hussman's mismatch index HMI (our notation), given by Equation 4:

4. $$HMI = \frac{market\ capitalization}{gross\ value\ added}$$

The two components of HMI are plotted in Figure 3. The solid thick series is market capitalization (the numerator of Equation 4), while the dashed series is gross value added (the denominator). Both series pertain to U.S.-based nonfinancial corporations, both are denominated in dollars, and each is plotted against a log scale.[13]

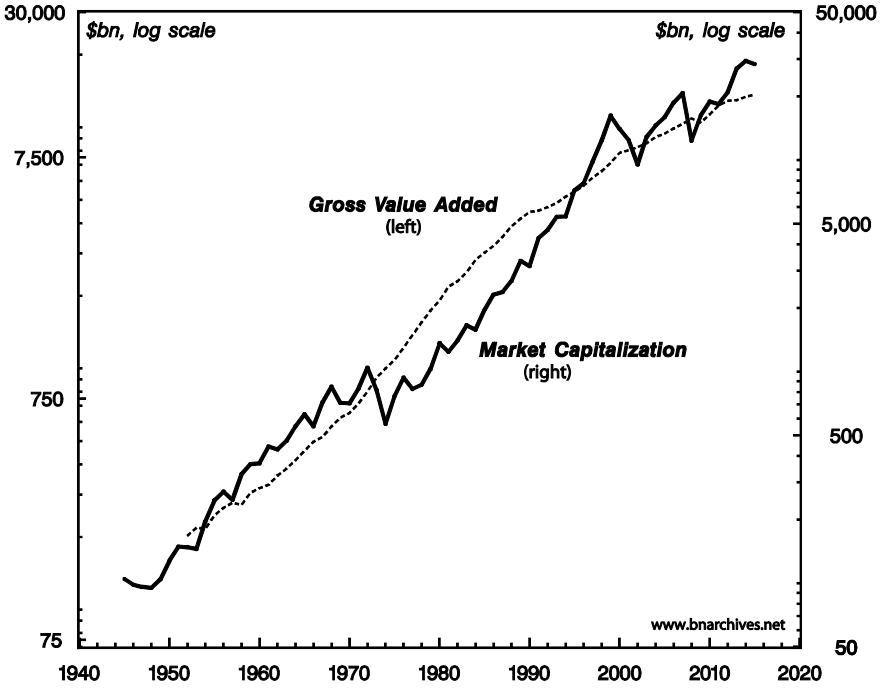
Figure 4, which reproduces Hussman's basic chart, plots Hussman's mismatch index HMI against the left log scale. Equation 5 shows that this indicator is equal to the four-element component from the right-hand side of Equation 3. And as the chart suggests, the index is indeed mean-reverting: it fluctuates roughly between a low of 0.6 and a high of 2.3.

5. $$HMI = \frac{K}{GVA} = \frac{m \times H}{nrr \times \delta}$$

[12] Hussman's own model uses monthly data, whereas our reproduction here employs annual data. The model's rationale, underlying assumptions and technical specifications are provided in Hussman (2015a, 2015b). Hussman's articles are archived here: http://www.hussmanfunds.com/weeklyMarketComment.html.

[13] Hussman's decision to focus on non-financial corporations rather than all corporations is not theoretical, but practical: the former cluster offers slightly better predictions than the latter.

Figure 3: U.S.-Based Nonfinancial Corporations



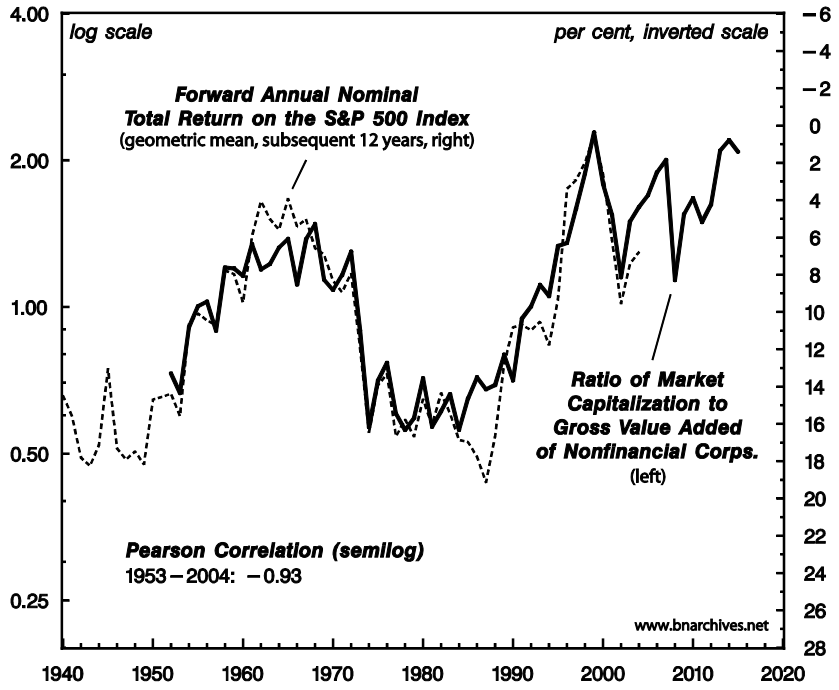
NOTE: Market value of nonfinancial corporations includes assets held domestically and in the rest of the world. Gross value added of nonfinancial corporations is domestic gross value added augmented by the imputed gross value added of foreign operations. It is computed by multiplying domestic gross value added of the nonfinancial corporate sector by (1 + U.S. profit from foreign operations/U.S. domestic after-tax profit), with both profit components smoothed as 5-year trailing averages. The last data points are for 2015.

SOURCE: Market value of nonfinancial corporations is from the Federal Reserve Board Flow of Funds through GI (series codes: LM103164103 for domestic assets and LM263164103 for assets held in the rest of the world). Domestic gross value added of nonfinancial corporations is from the Bureau of Economic Analysis (BEA) through GI (series code: GVANFC). U.S. after-tax profit is from the BEA through GI (series codes: ZAD for domestic after-tax profit, XFYADIV for foreign dividend income and XFYAREONUSDI for reinvested foreign earnings).

4.2 Predicting forward returns

Now, why should this mismatch ratio be of any interest to investors? According to Hussman, the reason is given by the dashed series in Figure 4, plotted against the inverted right scale. This series measures the forward annual nominal total rate of return on equities twelve years into the future. In other words, every observation in the series tells us what investors who had bought the S&P 500 index in that particular year ended up earning in capital gains and dividends, on average, over the next twelve years.

Figure 4: Hussman's U.S. Valuation index and forward returns



NOTE: Market value of nonfinancial corporations includes assets held domestically and in the rest of the world. Gross value added of nonfinancial corporations is domestic gross value added augmented by the imputed gross value added of foreign operations. It is computed by multiplying domestic gross value added of the nonfinancial corporate sector by (1 + U.S. profit from foreign operations/U.S. domestic after-tax profit), with both profit components smoothed as 5-year trailing averages. Forward annual nominal total return on the S&P 500 is calculated by (1) computing the ratio between the total return index 12 years ahead and its current value, and (2) taking the twelfth root of that ratio, subtracting 1 and multiplying by 100. The semilog correlation is between the log of the valuation index and the forward return. The last data points are 2015 for the ratio of market value to gross value added and 2004 for forward annual nominal total returns.

SOURCE: Nominal total return for the S&P 500 is from Global Financial Data (GFD) till 1969 (series code: _SPXTRD) spliced with data from Global Insight (GI) for 1970 onward (series code: SP500TRI). Market value of nonfinancial corporations is from the Federal Reserve Board Flow of Funds through GI (series codes: LM103164103 for domestic assets and LM263164103 for assets held in the rest of the world). Domestic gross value added of nonfinancial corporations is from the Bureau of Economic Analysis (BEA) through GI (series code: GVANFC). U.S. after-tax profit is from the BEA through GI (series codes: ZAD for domestic after-tax profit, XFYADIV for foreign dividend income and XFYAREONUSDI for reinvested foreign earnings).

For example, an investor who bought the S&P 500 index in 1984, when HMI was less than 0.6, ended up earning an annual average of 16 per cent over the next twelve years. And how do we know? We simply project the observation horizontally to the inverted right-hand scale and read the result. Or take an investor who bought the market at its 1999 peak, when HVI was nearly 2.3. Projecting the dashed series onto the right-hand scale shows that this investor ended up making an average nominal total return of less than 1 per cent over the next twelve years.

The key thing to note here is that forward total returns and the Hussman mismatch index are *nearly perfectly inversely correlated*: their Pearson correlation coefficient is –0.93 out of a maximum of –1. On the face of it, this result is nothing short of remarkable: while equity owners might find it difficult to predict what will happen to their investments in the next few

months or the next couple of years, according to Hussman's mismatch model in Figure 4 they can know pretty well what they will earn over the next twelve years. (At a risk of putting the cart before the horse, note that the current reading of Hussman's mismatch index is more or less the same as it was in 1999, just before the onset of the latest MBM. What this reading implies can be summarized in three ominous words: 'Winter is Coming'.)

4.3 Irrationality, risk and return

But the model's nearly perfect predictive record is somewhat tainted by two troubling observations. The first observation concerns 'economic rationality'. Mainstream economists insist that accumulation thrives on and fosters rationality (by which they mean that money prices reflect utility, and that this reflection implies Pareto-optimal-resource-allocation-read-utility-maximization). But Figure 4 seems to suggest the exact opposite: the more irrational the investors, the greater the apparent mismatch between nominal market capitalization and the real fundamentals; and the larger the mismatch, the greater the scope for accumulation – either by buying the market long when it is 'oversold', or shorting it when it is 'overbought' (relative to the fundamentals). In other words, utility-maximizing investors have an interest in – and therefore an incentive to foster – 'economic *ir*rationality'.

The second observation has to do with risk and return. Finance textbooks reiterate the Austrian economic mantra that the two go hand in hand: since there can be no gain without pain, or so they say, those who wish to earn more must be willing to take on a greater risk. But here too Figure 4 seems to imply otherwise: according to the historical data, the greatest future returns are available when risk is lowest (such as in the early 1980s), whereas when risk is very high (for instance, during the late 1990s), eventual returns are dismal. In other words, investors are either free riders who get something for nothing, or suckers who pay something for nothing. . . .[14]

5. The power model

Having outlined the conventional cosmology of finance, let us now turn it upside down by sketching our own, triangular CasP model of the stock market. We do so in three steps: in this section we examine the stock market from the viewpoint of capitalized power; in the next section we relate this capitalized power to systemic risk; and in the subsequent section we tie both capitalized power and systemic risk with strategic sabotage.

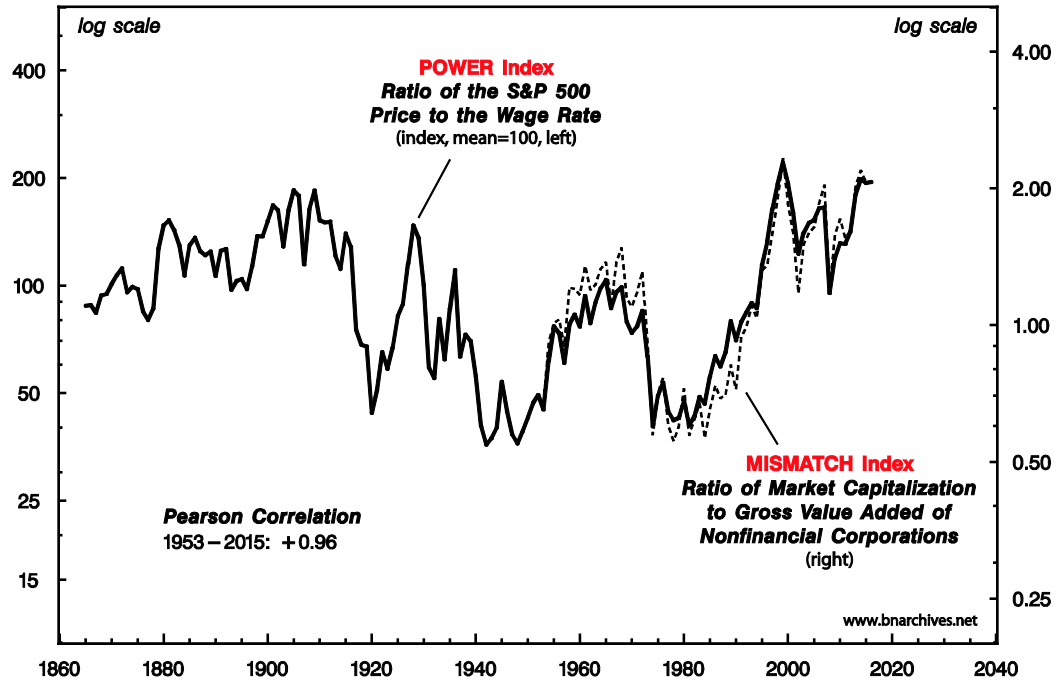
5.1 The power index (PI)

Figure 5 introduces our power index and contrasts it with Hussman's. The dashed series is Hussman's mismatch index HMI plotted against the right log scale. The solid series, plotted against the left log scale, is our own power index PI; it is defined as the ratio between the S&P 500 price index and the average wage rate, normalized with its historical mean=100:

6. $$PI = \frac{stock\ price\ index}{wage\ rate}$$

[14] Hussman's position on this issue is not entirely clear. On the one hand, he repeatedly alerts his readers that extremely overvalued stocks represent a toxic combination of maximum risk and minimum return (see, for instance, 2000, 2016). On the other hand, as far as we can tell, he keeps silent on why this theoretically awkward risk/return profile should exist in the first place.

Figure 5: U.S. Equity valuations: mismatch or power?



NOTE: Market value of nonfinancial corporations includes assets held domestically and in the rest of the world. Gross value added of nonfinancial corporations is domestic gross value added augmented by the imputed gross value added of foreign operations. It is computed by multiplying domestic gross value added of the nonfinancial corporate sector by (1 + U.S. profit from foreign operations/U.S. domestic after-tax profit), with both profit components smoothed as 5-year trailing averages. The S&P 500 price splices the following four sub-series: a combination of bank, insurance and railroad stock series weighed by Global Financial Data (1820-1870); the Cowles/Standard and Poor's Composite (1871-1925); the 90-stock Composite (1926-1956); and the S&P 500 (1957-present). The wage rate splices the hourly wage rate for manufacturing production workers till 1946 with the hourly wage rate for nonfarm business-sector workers from 1947 onward. The last data points are 2015 for the mismatch index and 2016 for the power index.

SOURCE: Market value of nonfinancial corporations is from the Federal Reserve Board Flow of Funds through GI (series codes: LM103164103 for domestic assets and LM263164103 for assets held in the rest of the world). Domestic gross value added of nonfinancial corporations is from the Bureau of Economic Analysis (BEA) through GI (series code: GVANFC). U.S. after-tax profit is from the BEA through GI (series codes: ZAD for domestic after-tax profit, XFYADIV for foreign dividend income and XFYAREONUSDI for reinvested foreign earnings). The S&P 500 price is from Global Financial Data (GFD) till 1900 (series code: _SPXD) and from Global Insight (GI) from 1901 onward (series code: JS&PNS). The hourly wage rate splices the following series: *Historical Statistics of the United States*, Millennial Edition Online: hourly wages in manufacturing, all trades, 1865-1889 (series code: Ba4290), hourly earnings in manufacturing, all industries, 1890-1913 (series code: Ba4299), weekly earnings of production workers in manufacturing, 1914-1918 (series code: Ba4362), hourly earnings of production workers in manufacturing, 1919-1938 (series code: Ba4361); Global Insight (GI): average hourly earnings of production workers in manufacturing, 1939-1946 (series code: AHPMFNS); Conference Board through GI: average hourly compensation of all employees in the nonfarm business sector (series code: JRWSSNFE).

Why is the ratio of the S&P 500 price index to the wage rate a 'power index'? To answer this question, recall that Hussman's mismatch index HMI tries to match market capitalization (which represents the aggregate purchasing power of equity holders) with the underlying fundamentals of this capitalization (a proxy for the aggregate income services to be produced by the corresponding real capital stock).

Our own power index PI is very different. First, it shifts the discussion from ontological to operational symbolism.[15] Note that, unlike Hussman, we compare not *overall* purchasing power and income services, but merely *prices*. For Hussman, nominal market capitalization is an ontological symbol: it represents, accurately or inaccurately, an external entity out there – in this case, the 'real' utils to be generated by the 'real' capital stock. In contrast, we see stock prices and the wage rate as generative, or operational symbols. Their ratio does not *represent* an outside reality, it *constitutes* that reality – in this case, the capitalized power of owners imposed on the underlying population.

The second difference is that the constituents of the power index PI – stock prices and the wage rate – are totally different creatures. Unlike the numerator and denominator of Hussman's mismatch model, they have no reason to match and therefore no reason to mismatch.

This is no coincidence. Whereas the liberal universe tries to harmonize its categories, the CasP cosmos pits them against each other. And indeed, instead of mirroring one another, the numerator and denominator of PI represent a *conflict*: the clash between those who own the capitalized means of power and those who are controlled by them. Note that we use the average wage rate here not as a measure of productivity or wellbeing, but as a benchmark against which to gauge the differential power of owners. Furthermore, although strictly speaking the wage rate pertains only to employed workers, its temporal movement approximates, however crudely, the changing conditions of the underlying population at large. Thus, when our power index PI rises, this means not that the market is distorted or that investors are economically irrational, but that the power of equity owners relative to the underlying population increases – and vice versa when the index falls. Moreover, and importantly, this relative power is *forward looking*: it denotes not only the rulers' relative position here and now, but also how they expect this relative position to change in the future.

All in all, then, the mismatch and power indices are conceptually distinct and theoretically unrelated. And yet – and here we come to the important bit – despite these fundamental differences they correlate almost perfectly: according to Figure 5, their Pearson coefficient, extended over more than half a century, is +0.96 out of a maximum of +1.

5.2 Mismatch or power?

What explains this remarkably tight correlation? Is this a miraculous statistical fluke, or is there a hidden connection between these seemingly different indicators? The answer emerges from Equations 7-11, which relate Hussman's mismatch index HMI to our own power index PI. Equation 7 decomposes both market capitalization K and gross value added GVA to their respective price and quantity components (the average price of stocks P and the number of outstanding stocks N for the former, the gross value added deflator $GVAD$ and real gross value added Q for the latter).

[15] The notion of 'capital as power' as an operational symbol was first suggested and elaborated by Ulf Martin (2010). Following Sybille Krämer, Martin distinguishes between three types of symbols: magical (the entity is the symbol), ontological (the symbol represents a distinct worldly entity), and operational (the symbol defines the entity). According to Martin, capital as power, because it defines its own logic as well as creates and recreates the social reality, is best understood as an operational symbol. This 'generative/operational' property is highlighted by Cochrane (2016) who connects the changing structure of the Chinese family with the differential capitalization of the commodities-driven TSX relative to the S&P 500, and by Malik and Phillips (2012) and Malik (2014) who use the art market and derivatives to demonstrate the non-ontological nature of capitalization.

GVA	gross value added ($)
$GVAD$	gross value added deflator ($)
K	market capitalization ($)
N	number of shares outstanding
P	stock market price index ($)
$Price.Ratio$	$W/GVAD$ (decimal)
Q	'real' gross value added (decimal)
$Quantity.Ratio$	N/Q (decimal)
$Residual$	$N/Q \times W/GVAD$ (decimal)
W	wage rate ($)

7. $$HMI = \frac{K}{GVA} = \frac{P \times N}{GVAD \times Q}$$

Equation 8 divides and multiplies the denominator of Hussman's mismatch index by the wage rate W and rearranges the terms:

8. $$HMI = \frac{P \times N}{\frac{GVAD}{W} \times W \times Q} = \frac{P}{W} \times \frac{N}{Q} \times \frac{W}{GVAD}$$

Equations 9-11 show further rearrangements, to accentuate the connection between the mismatch and power indices:

9. $$HMI = PI \times \frac{N}{Q} \times \frac{W}{GVAD}$$

10. $$HMI = PI \times Quanity.Ratio \times Price.Ratio$$

11. $$HMI = PI \times Residual$$

Equation 11 shows that the power index PI – i.e. the capitalized conflict between equity owners and the underlying population – is in fact part and parcel of Hussman's mismatch index HMI, while the nearly perfect correlation in Figure 5 suggests the former index is the main driver of the latter. The combined impact on HMI of the quantity and price ratios is a negligible residual.[16]

In this sense, the mismatch theory reminds us of the Ptolemaic geocentric view. The idea that the sun epicycles around the earth yielded fairly accurate predictions, but it was nonetheless wrong.[17] And perhaps the same might be said about the notion that capitalists and analysts price the stock market to match the so-called real fundamentals of utility and productivity: just like the geocentric view, this notion yields very tight predictions, and just like the geocentric view, it seems completely misplaced. The real driving force here is not the mismatching of future utility, but changes in organized power.

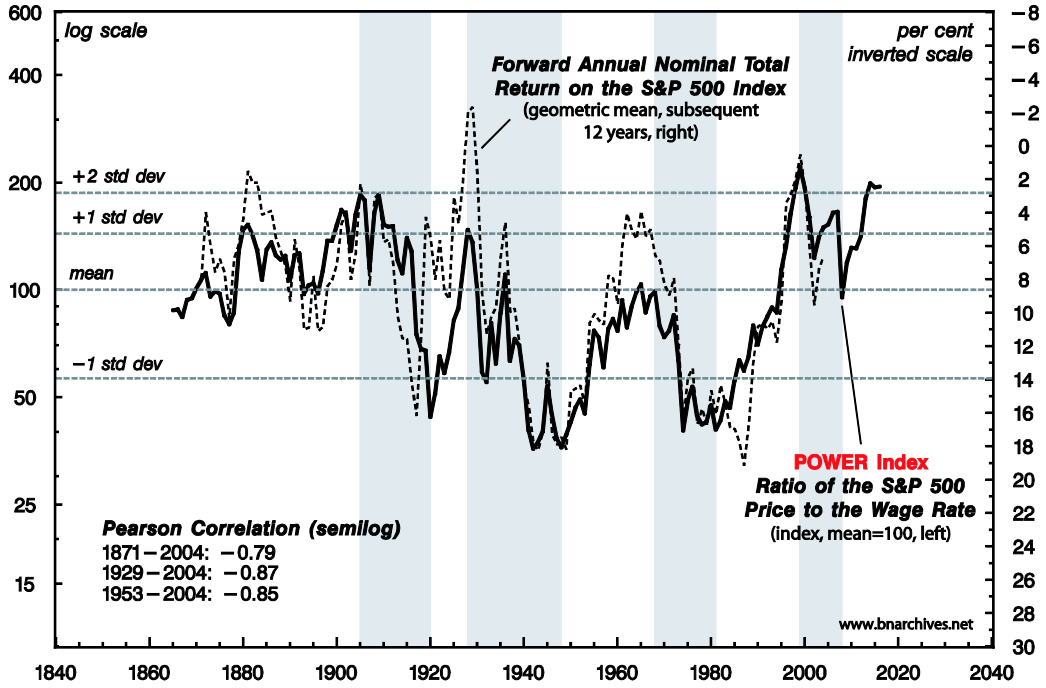
[16] Historically, the $Quantity.Ratio$ has trended downward while the $Price.Ratio$ has trended upward. Moreover, the two movements have more or less offset each other, so their product – the $Residual$ in Equation 11 – has ended up moving sideways, showing little or no correlation with the mismatch and power indices.

[17] For a succinct comparison between the old geocentric model and its heliocentric alternative, see Singh (2004, particularly Table 2, pp. 34-35).

5.3 Capitalized power and forward returns

Moreover, according to Figure 6, it appears that it is *this actual reality* of capitalized power – and not the *mismatched reality* of utility and productivity – that drives future returns. The chart displays two series. The solid series is our power index, plotted on the left log scale. The dashed series, plotted on the right inverted scale, is the forward annual nominal total returns on the S&P 500, projected twelve years ahead. For convenience, the chart shows the historical mean of the power index (=100), one standard deviation below and two standard deviations above it and the last four MBMs (shaded).

Figure 6: The power index and forward returns



NOTE: Shaded years denote major bear markets (MBMs) as defined in Table 1. Series are normalized with their historical mean=100. The S&P 500 price splices the following four sub-series: a combination of bank, insurance and railroad stock series weighed by Global Financial Data (1820-1870); the Cowles/Standard and Poor's Composite (1871-1925); the 90-stock Composite (1926-1956); and the S&P 500 (1957-present). The wage rate splices hourly data for manufacturing production workers till 1946 with hourly data for nonfarm business-sector workers from 1947 onward. Forward annual nominal total return on the S&P 500 is calculated by (1) computing the ratio between the total return index 12 years ahead and its current value, and (2) taking the twelfth root of that ratio, subtracting 1 and multiplying by 100. The semilog correlation is between the log of the power index and the forward return. The last data points are 2015 for the power index and 2004 for forward annual nominal total returns

SOURCE: The S&P 500 price is from Global Financial Data (GFD) till 1900 (series code: _SPXD) and from Global Insight (GI) from 1901 onward (series code: JS&PNS). The hourly wage rate splices the following series: *Historical Statistics of the United States*, Millennial Edition Online: hourly wages in manufacturing, all trades, 1865-1889 (series code: Ba4290), hourly earnings in manufacturing, all industries, 1890-1913 (series code: Ba4299), weekly earnings of production workers in manufacturing, 1914-1918 (series code: Ba4362), hourly earnings of production workers in manufacturing, 1919-1938 (series code: Ba4361); Global Insight (GI): average hourly earnings of production workers in manufacturing, 1939-1946, (series code: AHPMFNS); Conference Board through GI: average hourly compensation of all employees in the nonfarm business sector (series code: JRWSSNFE). Nominal total return for the S&P 500 is from GFD till 1969 (series code: _SPXTRD) spliced with data from Global Insight (GI) for 1970 onward (series code: SP500TRI).

The bottom left of the chart lists Pearson correlation coefficients for different periods. In general, the correlation between capitalized power and future returns is weaker before 1929 than after. The relevant period for comparison, though, is 1953-2004 – the years covered by Hussman's model in Figure 4. For this period, the correlation in Figure 6 is –0.85, only a bit lower than Hussman's –0.93 (which is to be expected, given the additional contribution of the $Residual$ factor).

5.4 Irrationality, risk and return (reconsidered)

So the difference between the mismatch and power indices is not empirical but theoretical, and here we should return to the two points raised at the end of the previous section: the issue of irrationality and the relationship between risk and return.

Begin with irrationality. If capitalists and analysts were perfectly rational and fully prescient, the four-component element on the right-had side of Equation 3 would be a fixed number, market capitalization would fluctuate in tandem with gross value added and Hussman's mismatch index would trace a straight horizontal line equal to the index's historical mean. This is what a perfectly matched world *should* look like.

But that is not what we see in practice. Looking at the power index in Figure 6 – which, as we have seen in Figure 5, is nearly a carbon copy of Hussman's mismatch index – we see that fewer than 15 per cent of the observations are equal to the index's normalized mean of 100. The rest are either bigger or smaller. In other words, from the viewpoint of the mismatch thesis, the market is, by and large, economically *irrational*. Moreover – and here we come to the key point – the irrationality is almost always blamed on *various forms of power*, from asymmetric information and policy mistakes to market imperfections and extra-economic interventions. But then, if valuation is almost always out of tune and its deviations are mostly a matter of power, why not put aside irrationality and distortions and focus directly on what everyone seems to agree matters the most – namely power? Once the focus shifts to power, there is nothing to match and therefore no mismatch. And with the mismatch gone, there is no longer anything irrational about the valuation index going up and down. Its movements simply reflect the changing landscape of power.

The second difference has to do with risk and return. As noted, the annals of finance stipulate that risk and return should be positively correlated – yet, according to both the mismatch and power models, their correlation is in fact negative. Now, while this negative correlation does not sit well with the economic underpinnings of the mismatch thesis, it is perfectly consistent with the power underpinnings of our CasP model. Just like the mismatch index, capitalized power predicts forward returns *negatively*: the higher the power, the lower the return. And the reason for this negative relation is twofold: (1) power is always exercised through some form of strategic sabotage (a concept to which we return below), and sabotage elicits resistance; and (2) the greater the power and sabotage, the greater the provoked resistance. Everything else being the same, it is easier for capitalists to augment their power when the power index is at one or two standard deviations below its average (like it was in the 1940s and 1980s, for example) than when the index is one or two standard deviations above it (like in the 1900s, the 1990s and now). It is this changing intensity of resistance – and the fact that resistance goes hand in hand with power – that makes our power index mean reverting, and it is this mean reversion that ascertains that risk and return will be related negatively rather than positively.

In this respect, it seems that two standard deviations above the mean is the historical asymptote, or limit, of capitalized equity power in the United States.[18] This asymptote was reached three times over the past 150 years – in the early 1900s, in the late 1990s, and now – and in the previous two times this has happened, there followed a significant reversal in the form of an MBM. In other words, when capitalized power approaches its asymptotes, capitalists have good reason to fear their *very own power*. And here we come to the second aspect of our model: the phenomenon of systemic fear.

6. Systemic fear

Seven years ago, in the midst of the financial crisis, we wrote a paper in which we argued that capitalists were struck by systemic fear: that they were apprehensive not about rising interest rates or falling profit, but about the *very existence* of the system as they knew it.[19] We also argued that their systemic fear could be identified empirically, by looking at the co-movement of stock prices and corporate profit (Nitzan and Bichler 2009b; expanded in Bichler and Nitzan 2010).

6.1 Identifying systemic fear

Our logic could be summarized as follows. Capitalization is forward-looking: according to valuation guru Benjamin Graham, it should discount not current profit, but the profits that will be earned in the future, all the way to 'eternity' (quoted in Zweig 2009: 28). In this scheme, *current variations* in profit have no more than a negligible impact on the final outcome. And indeed, if you revisit Equation 3, you will see that capitalization in this equation depends on the coefficient m, which in turn hinges on the average future share of profit in GVA and the average future growth rate of GVA. Current profit does not even appear there.

This ritualistic reliance on the future implies systemic confidence. It demonstrates a belief that earnings will continue to flow and that assets will always have buyers – in other words, that the system is eternal, and that the operational symbol of capitalization will dominate the world forever.

Now, imagine the very opposite situation – a setting in which capitalists lose this systemic confidence in the future and are instead struck by systemic fear. What happens when they start to hesitate? What happens when the power index is at an all-time high and capitalists become concerned that the current power architecture is unsustainable? When the chief promoters of globalization throw in the towel, saying that globalization doesn't work? When some peripheral states drift out of their superpower orbit while others disintegrate altogether? When conventional economic predictions fail miserably and domestic policymakers seem clueless? When inequality reaches historical extremes and the underlying population simmers with discontent? When the ecosystem destabilizes and resource extraction faces exhaustion?

[18] For different analyses of the asymptotes of power, see Bichler and Nitzan (2012a), Kliman, Bichler and Nitzan (2011) and Bichler and Nitzan (2014).
[19] 'For this consciousness [of the capitalist bound to the steering wheel of a megamachine gone wild] was not in peril and fear for this element or that [such as falling profit or rising volatility], nor for this or that moment of time [like a sharp market correction or a declaration of war], it was afraid for its entire being; it felt the fear of death, the sovereign master [the ultimate wrath of the ruled]. It has been in that experience melted to its inmost soul, has trembled throughout its every fibre, and all that was fixed and steadfast has quaked within it [will capitalism survive?]' (Hegel 1807: 237; paraphrased in Bichler and Nitzan 2010: 19).

The result of these developments is systemic fear, an apprehension that the current mode of power might crumble.

The rulers' immediate reaction to systemic fear, though, is not capitulation, but *denial*: 'What? We, capitalists, worry? Fear for our system? No way!' But, then, to sustain this denial and retain a semblance of confidence, capitalists need evidence that they are still very much in driver's seat, and the most readily available evidence of such control is current profit. If current profit remains high – or better still, if it continues to rise – then we, the capitalists, can remain hopeful despite the threatening future. And if our group as a whole stays hopeful, then, as individual investors, we all have good reason to hold on to and even augment our equity portfolios.

Paradoxically, then, the evidence for systemic fear lies in its very denial. We can know that capitalists *have been* struck by systemic fear by the fact that they effectively negate and abandon their core ritual of forward-looking capitalization; and we can know the *degree* to which they negate this ritual by the extent to which their asset pricing comes to depend on current rather than future earnings.

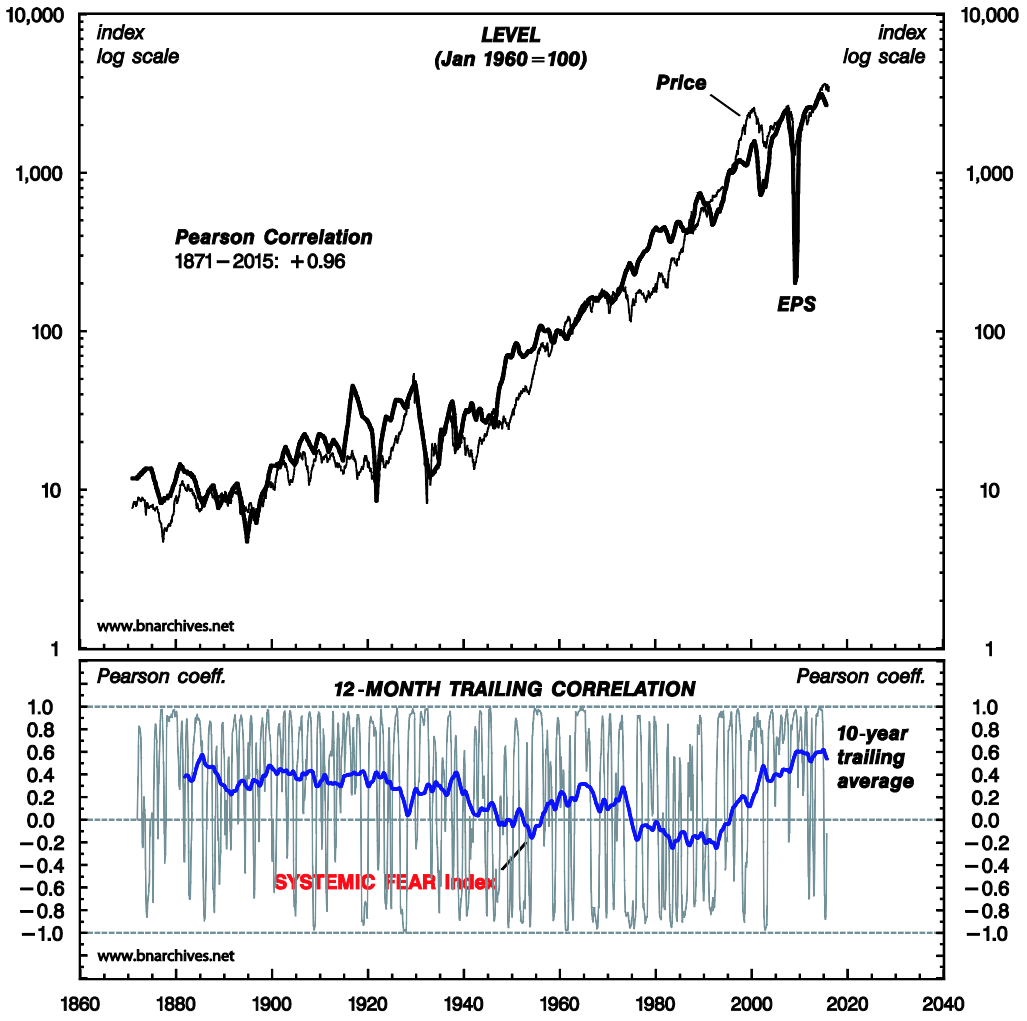
In our earlier work we used this empirical proof-by-negation to argue that there were two periods during which capitalists were struck by systemic fear: the 1930s and the 2000s (Nitzan and Bichler 2009b; Bichler and Nitzan 2010). We demonstrated our argument in three steps: first by plotting the annual rates of change of equity prices and current earnings per share (EPS) smoothed as three-year trailing averages; then by visually comparing the co-movements of these rates of change; and finally by identifying the 1930s and the 2000s as the only two periods in which the two rate-of-change series seemed tightly correlated.[20] Unfortunately, though, our method was not very rigorous, and that lack of rigor caused us to make a serious empirical mistake: we failed to identify two additional periods in which the series were positively correlated, and this failure incited a heated debate in the *Journal of Critical Globalization Studies* (Kliman, Bichler, and Nitzan 2011).

6.2 The systemic fear index

So here we try to do things differently. First, we look not at the rates of change of stock prices and EPS, as we originally did, but at their *levels*. Measures of correlation already reflect change, so to correlate rates of change is akin to looking at the second instead of the first derivative. Second, we *shorten* the window of comparison between price and current EPS. Our original three-year window was probably far too long for short-sighted capitalists, so here we reduce it to twelve months. Third and finally, instead of merely eyeballing the correlation window between the rates of change of stock prices and EPS, we *measure it systematically* as it drifts over the entire dataset.[21]

[20] Strictly speaking, the term 'current' earnings per share is a misnomer: in practice, EPS data are reported as the average earnings per share recorded over the *previous* twelve months. In this sense, all current earning data are in fact *backward*-looking.
[21] The use of a moving correlation here was suggested and empirically demonstrated by Ulf Martin (private communication with the authors, May 2014).

Figure 7: The S&P 500 price and EPS: the systemic fear index



NOTE: The S&P 500 index splices the following four sub-series: a combination of bank, insurance and railroad stock series weighed by Global Financial Data (1820-1870); the Cowles/Standard and Poor's Composite (1871-1925); the 90-stock Composite (1926-1956); and the S&P 500 (1957-present). The 12-month trailing correlation in the bottom panel (thin series) measures the correlation between price and EPS (earnings per share). The ten-year trailing average (thick series) is the mean of this trailing correlation over the past 120 months. The last data points are September 2015 for EPS and April 2016 for price.

SOURCE: S&P 500 price and EPS are from Robert J. Shiller's online data archives (http://www.econ.yale.edu/~shiller/data/ie_data.xls, accessed on April 20, 2016).

The result of these modifications is displayed in Figure 7. The top panel of the chart shows normalized monthly price and EPS data for the S&P 500 group of companies, dating back to 1871. The bottom panel plots short-term correlations. The thin series in the bottom panel measures the 12-month trailing correlation between the price and EPS series shown in the top panel. Each observation shows the correlation over the past year, with a value ranging between –1 (perfect inverse correlation) and +1 (perfect direct co-movement).

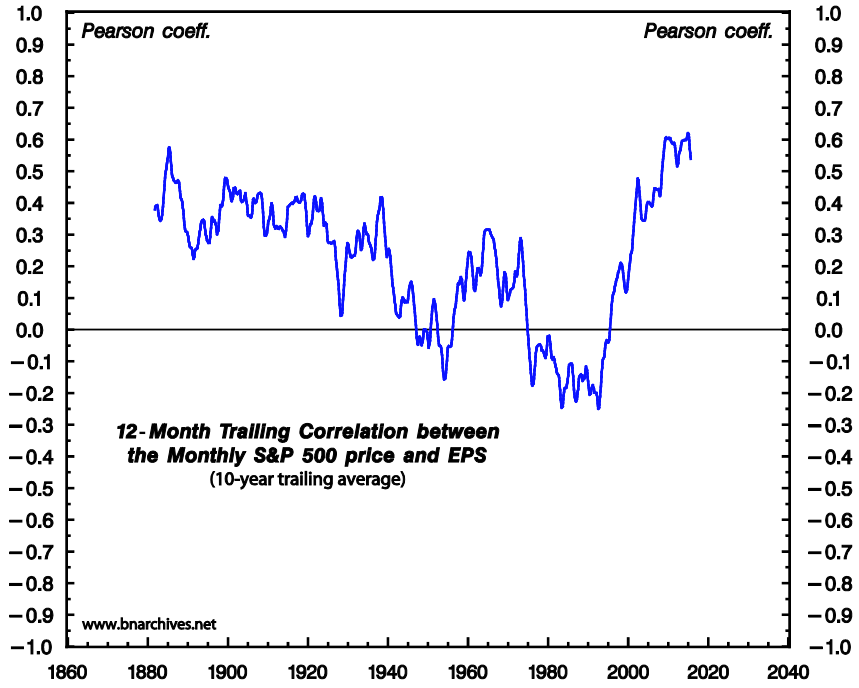
The difficulty with the thin 12-month trailing correlation is that it oscillates widely, so visual inspection alone is not very revealing here. The thick series in the bottom panel addresses

this difficulty by smoothing the thin series as a ten-year trailing average. Each observation in the thick series measures the average 12-month trailing correlation between price and EPS over the previous ten years. We call this series the *systemic fear index*.[22]

6.3 The historical evolution of systemic fear

Figure 8 enlarges our systemic fear index taken from the bottom panel of Figure 7, making it easier to examine.

Figure 8: The Systemic Fear Index



NOTE: The systemic fear index is calculated in two steps: (1) computing the 12-month trailing correlation between price and EPS (earnings per share) of the S&P 500 index; (2) calculating the ten-year trailing average of the 12-month moving correlation computed in the first step. The S&P 500 index splices the following four sub-series: a combination of bank, insurance and railroad stock series weighed by Global Financial Data (1820-1870); the Cowles/Standard and Poor's Composite (1871-1925); the 90-stock Composite (1926-1956); and the S&P 500 (1957-present). The last data point is for September 2015.

SOURCE: S&P 500 price and EPS are from Robert J. Shiller's online data archives (http://www.econ.yale.edu/~shiller/data/ie_data.xls, accessed on April 20, 2016).

The chart shows two clear patterns: one long term, the other shorter term. The long-term pattern has a V-shape, with the early 1990s as its low point. Until the early 1920s, forward-looking capitalization was still in its infancy, so the correlation between price and EPS was pretty high, hovering around +0.4. But even then there was already a visible down drift, and by the early 1940s this down drift had turned into a sharp decline. Discounting methods were now making their way into introductory textbooks, and by the 1950s, with the capitalization ritual becoming more widely accepted and increasingly internalized by equity investors, the

[22] For example, the systemic fear index for September 2015 (the last observation in the series) is +0.538. This result is derived by averaging out the 120 monthly readings of the 12-month trailing correlations between August 2005 and September 2015.

correlation fell to around zero and from then onwards continued to hover around this value – albeit with some significant oscillations. The decisive reversal came only in the early 1990s. Initially, the uptick looked like part of yet another short oscillation. But by the early 2000s it became evident (at least in retrospect) that the century-long downtrend had been broken. Instead of reverting back to zero, the systemic fear index continued to soar and, by the early 2010s, reached an all-time high of +0.6.

This V-shape pattern, though, has been anything but smooth. Oscillating around the long-term down- and uptrends we can see plenty of shorter-term fluctuations, some of which are pretty pronounced. So the question we need to address is what lies behind these patterns: what determines the long-term V-shape of the index and what accounts for its shorter-term fluctuations?

6.4 Culture or power?

To the best of our knowledge, this question has never been asked, let alone answered. Indeed, as far as we know, the V-shape pattern of the short-term price-EPS correlation shown in Figures 7 and 8 is a new finding.

It is common to argue that, since the 1980s, U.S. capitalism has been marked by a growing emphasis on 'shareholder value', heightened 'short-termism' and a nearly universal obsession with quarterly increases in profit. This popular view is certainly consistent with the post-1980s surge of the price-EPS correlation shown in Figure 8 – and this consistency should hardly surprise us. With capitalists paying more and more attention to the latest bottom line and analysts glued to the latest bit of news, it is no wonder that equity markets have become increasingly sensitive to the most recent variations in earnings.[23]

But what is the *cause* of these changes? Why has the capitalist time horizon shrunk? Why have investors who, for a whole century up until that point, cared less and less about current earnings and often seemed perfectly happy to buy and hold stocks for the long haul, suddenly started to insist on quarterly increases in profit? Is the V-shape reversal of the early 1990s merely the consequence of a changing 'investment culture'? Is it simply a new fad imprinted by the theoretical winds of just-in-time neoliberalism and emboldened by the ideological flare of Margaret Thatcher, Ronald Reagan and Alan Greenspan – or are these developments themselves the result of a deeper change?

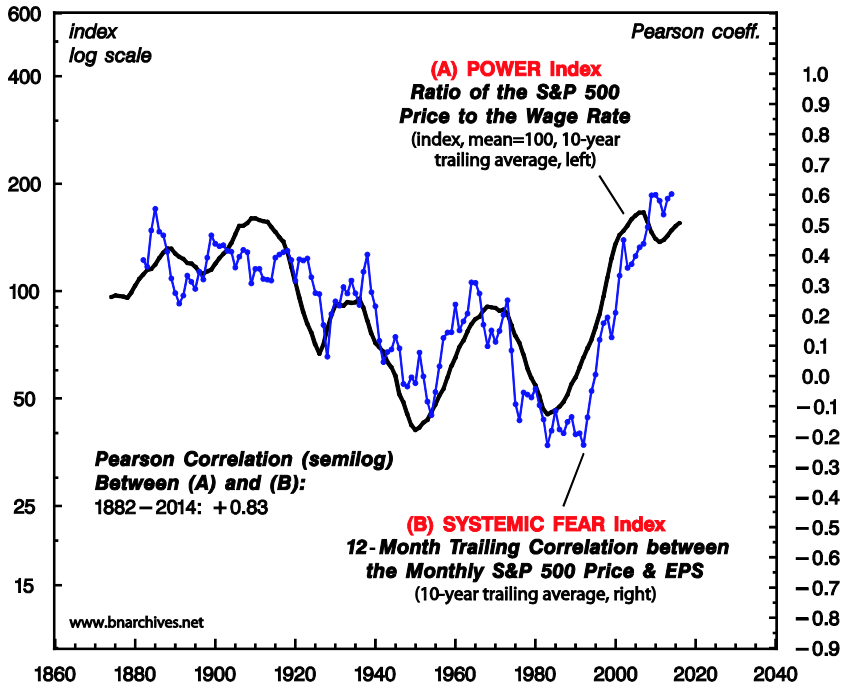
The evidence presented below suggests the latter. Present-day capitalists and analysts, we argue, have come to demand quarterly increases in profits not because they started to 'feel like it', because they were taken over by a new financial 'fashion' or because they were somehow convinced that short-term increases are more 'economically efficient' than long-term growth. In our view, they do so because they are *compelled* to, and the force that compels them has nothing to do with any of the above. The reason, rather, is that their capitalized power is approaching its asymptotes, and the only way for them to counteract their deepening systemic fear is by pushing for higher current earnings.

[23] This point was raised by Suhail Malik at the 2016 CasP conference presentation of this paper (http://bnarchives.yorku.ca/489/).

6.5 The co-movement of capitalized power and systemic fear

The long-term relationship between systemic fear and capitalized power is shown in Figure 9. The chart displays two series. The dotted blue series, plotted against the right scale, is our systemic fear index, taken from Figure 8. To reiterate, this index is the ten-year trailing average of the 12-month trailing correlation between the S&P 500 price and EPS. The solid black series, plotted against the left log scale, is our power index, which we take from Figure 6 and smooth as a ten-year trailing average to match the periodicity of the systemic fear index.

Figure 9: **The dialectic of power and fear**



NOTE: The systemic fear index represents annual averages of the monthly series shown in Figure 7 (see definition there). The S&P 500 price splices the following four sub-series: a combination of bank, insurance and railroad stock series weighed by Global Financial Data (1820-1870); the Cowles/Standard and Poor's Composite (1871-1925); the 90-stock Composite (1926-1956); and the S&P 500 (1957-present). The wage rate splices hourly data for manufacturing production workers till 1946 with hourly data for nonfarm business-sector workers from 1947 onward. The last data points are 2014 for the systemic fear index and 2016 for the power index.

SOURCE: Annual S&P 500 price is from Global Financial Data (GFD) till 1900 (series code: _SPXD) and from Global Insight (GI) from 1901 onward (series code: JS&PNS). The hourly wage rate splices the following series: *Historical Statistics of the United States*, Millennial Edition Online: hourly wages in manufacturing, all trades, 1865-1889 (series code: Ba4290), hourly earnings in manufacturing, all industries, 1890-1913 (series code: Ba4299), weekly earnings of production workers in manufacturing, 1914-1918 (series code: Ba4362), hourly earnings of production workers in manufacturing, 1919-1938 (series code: Ba4361); Global Insight (GI): average hourly earnings of production workers in manufacturing, 1939-1946 (series code: AHPMFNS); Conference Board through GI: average hourly compensation of all employees in the nonfarm business sector (series code: JRWSSNFE). Monthly S&P 500 price and EPS are from Robert J. Shiller's online data archives (http://www.econ.yale.edu/~shiller/data/ie_data.xls, accessed on April 20, 2016).

The correlation between the two series is extremely tight: its Pearson coefficient for the past 132 years is +0.83 out of a maximum of +1. What this correlation tells us is that the greater the capitalized power of equity owners relative to the underlying population, the greater their

systemic fear and therefore the greater their reliance on current earnings when pricing their stocks – and conversely, the lesser their capitalized power, the lower their systemic fear and hence the weaker their emphasis on present profit.

6.6 The dialectic of power and fear

At first sight, this co-movement might seem counterintuitive. Why should capitalists fear more for their system as they grow more powerful? Shouldn't it be the other way around – i.e., the greater their power, the *lesser* their systemic fear?

To answer this question, we need to backtrack a bit. Power is a complex and often slippery concept. It has numerous dimensions and layers, it is historically contingent and context-dependent and, most importantly, it is deeply dialectical and self-transformative. In our own research, we extend Johannes Kepler's scientific notion of force to view capitalized power not as a stand-alone qualitative entity, but as a *quantitative relationship between entities* (Nitzan and Bichler 2014: 141). In the present paper, we define this power very broadly as the relationship between equity owners and the underlying population, quantified by the ratio of stock prices to the wage rate. But we also argue that the quantity of capitalized power expresses the *rulers' confidence in the obedience of the ruled* (Nitzan and Bichler 2009a: 17) – which in our case here denotes the confidence of equity owners in the obedience of the underlying population.

Confidence in obedience, though, is not a monolithic sentiment. If we are to generalize, we might say that the buildup of power generates not one, but *two* movements – one extroverted, the other introverted – and that the trajectories of these two movements are not similar but *opposite*. On the outside, the relationship appears positive: the greater the rulers' power, the greater their display of confidence in obedience. But on the inside, the connection is negative: the more powerful the rulers, the greater their fear that their power might crumble.

This double-sided relationship is the linchpin of Hobbes' *Leviathan* (1691). The relatively equal abilities of human beings, he says, breed their uncertainty, insecurity and mutual suspicion, and these forces in turn compel them to try to increase their differential power without end. But, then – and this is the crucial qualifier – the more power one possesses, the more he or she dreads losing it all. The result is an ongoing cycle, with fear stoking a hunger for power, and the amassment of power heightening the very fear that begot that hunger in the first place (for example, pp. 75 and 94).

Now consider how this double movement unfolds in our case here. Capitalists, we posit, are driven to increase their capitalized power without end, and this increase, we maintain, boosts their expressed confidence in obedience. And how do we know that their confidence in obedience is indeed rising? Because the stock prices comprising the numerator of the power index are *determined by the capitalists themselves*, and because capitalists determine those prices by risking the thing they cherish the most: *their own money*. Indeed, the only reason for capitalists to buy stocks and in so doing bid up the stock price/wage ratio is that they expect this ratio to rise even further. And the fact that they believe that this ratio will go up attests to their confidence in obedience – the confidence that the underlying population will not expropriate them and that the system as a whole will not fail them. In this sense, our power index offers an *objective* measure of capitalist confidence – at least on the outside.

But as Figure 9 shows, there is another, inner process at work here: the *temporal basis* for capitalist confidence in obedience varies with the level of capitalized power. When the power index is low, the projected confidence of capitalists is inherently *forward-looking*. During such periods – for example, the 1940s or the 1980s – capitalists focus on the future and ignore present profit altogether (as indicated by the low, zero or even negative price-EPS correlation). And why? Because the lower the capitalized power, the *greater the scope for increasing it further*: income can be further redistributed in favour of profit, hype can be further amplified, profit volatility can be further decreased and the normal rate of return can be further lowered. And as long as these elements can be further augmented/reduced in favour of capital, owners can safely ignore the dismal present and focus on the promising future.

However, when the power index is high – as it was, for example, during the early twentieth century, and as it is now, at the beginning of the twenty-first – confidence in obedience has to rely largely on the *present* (and it does – as indicated by the high price-EPS correlation during these periods). And why? Because capitalized power is not unbounded. The greater the power, the greater the resistance to power. And when power approaches its asymptotes – in this case, when the profit share of income and the level of hype are already high and income volatility and the normal rate of return already low – increasing it further within the existing confines of the 'symbolic machine', as Ulf Martin (2010) calls it, becomes harder and harder.[24] Such increases require further threat, sabotage and open force, which in turn make the system ever more complex and increasingly brittle, and hence prone to breakdown (Bichler and Nitzan 2010). Under these circumstances, the only way for capitalists to retain their apparent confidence is to be constantly reassured that the system still holds here and now. And since the future is too bleak to rely on, this reassurance can come only from current profit.

6.7 The omen

Rulers always need an omen, a self-serving looking glass to bolster their confidence and galvanize their resolve. But sometimes the omen refuses to cooperate, and when it disobeys, the façade crumbles and the rulers find themselves facing the void. Literature offers many illustrious examples: the evil queen in the Brothers Grimm's *Little Snow-White*, whose obedient magic mirror suddenly defies her, declaring that she is not the fairest of all; Genghis Khan in Aitmatov's *The Day Lasts More than a Hundred Years* (1983), whose loyal guiding cloud suddenly disappears, leaving the Khan's globetrotting conquest in tatters; Belshazzar, the omnipotent king of Babylon, whose hubris is suddenly deflated by a mysterious writing on the wall (Book of Daniel: Ch. 5); the list goes on.

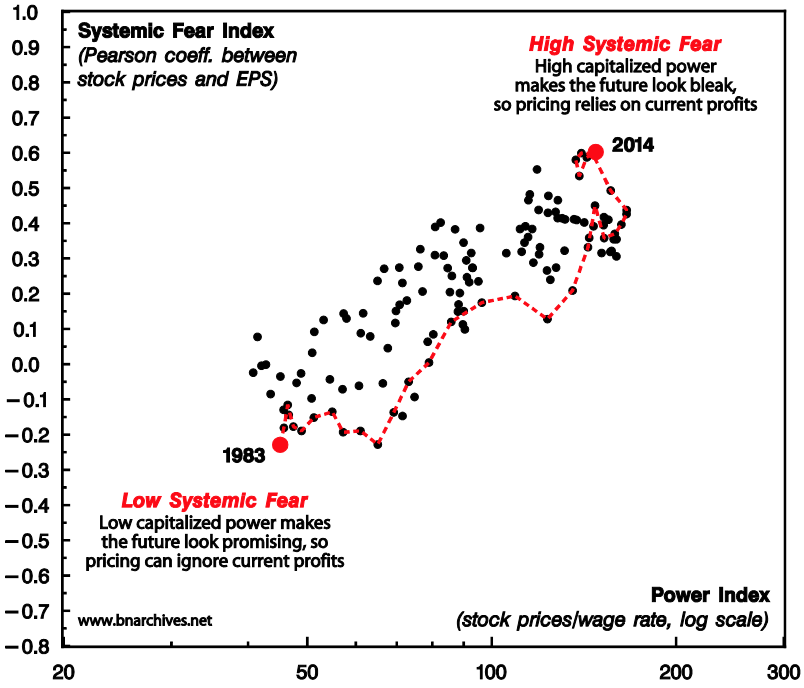
These power mirrors, though, are pretty naïve. They typically generate no more than a binary image, and their warnings almost always come too late. By contrast, the stock price-EPS correlation offers an infinitely nuanced reflection. Instead of a binary image, it draws a continuous scale, ranging from a Pearson coefficient of 0 (or less), which indicates that forward-looking capitalists do not fear for their system, to a Pearson coefficient of +1, which means that capitalists, struck by systemic fear, have abandoned their core belief in forward-looking capitalization in favour of a defensive, backward-looking posture.

This analytical range is shown historically in Figure 10. The chart presents the same data series from Figure 9, but instead of displaying them on a time scale, it plots them against one

[24] Bichler and Nitzan (2012a) examine some of these limits in the United States.

another. Each annual observation projects two readings: the ten-year trailing average of the power index on the horizontal scale, and the systemic fear index on the vertical scale. The observations are tightly clustered around a positive slope, reconfirming what we have already seen in Figure 9 – namely, that capitalized power is closely intertwined with systemic fear. For illustration purposes, we use a dashed red line to trace the evolution of this temporal relationship during the most recent period: from 1983, when the systemic fear index was at a record low, to 2014, when it reached its all-time high.

Figure 10: The dialectic of power and fear, 1882-2014



NOTE & SOURCE: See Figure 9.

The gradual temporal 'stretching' of this dashed line has been akin to pulling a string: as the United States moved up and to the right on this path, the tension between sabotage and resistance kept rising and rising. However, because the process has been so slow and drawn out, initially this buildup was largely imperceptible. Indeed, until recently the key 'actors' themselves – i.e., the capitalists and fund managers, along with policymakers and opinion shapers – remained largely unware of it and seldom admitted it, not even to themselves (and rarely if ever in the manner described here). But as Veblen might have put it, although they are yet to recognize it with their mind, they already know it in their heart. And here their *actions* speak louder than words: with their power rising, they have gradually but systematically abandoned their sacred ritual of forward-looking capitalization in favour of the still-rosy present. Their current mode of power is becoming increasingly unstable, and their short-term equity pricing indicates that underneath the hubris lies a deepening apprehension that it might not last.

Our own study of redistribution as the key power axis of capitalism started during the early 1980s. At the time, U.S. capitalized power and systemic fear were at all-time lows, investors were totally oblivious to the issue and our work was typically classified as 'social economics' (with an aftertaste of moralizing 'social justice'). But as capitalized power and systemic fear

increased, the crucial importance of redistribution slowly percolated to the surface, and in 2014, when power and fear reached record highs, Thomas Piketty's work on inequality (Piketty 2014) was suddenly made top news and everyone suddenly knew (all along) that the top 1 per cent held the rest of the world under its thumb.

And then the discourse started to change. Although the talking heads still hail capitalism as the best of all possible worlds, we now begin to see more and more expressions of guilt (the IMF admitting that the project of neoliberal globalization has been 'oversold'; Ostry, Loungani, and Furceri 2016), remorse (McKinsey cautioning that the current generation is poorer than its parents; McKinsey & Company *et al.* 2016) and dire warnings (former bond king Bill Gross alerting his fellow capitalists that, although 'I'm an investor that ultimately does believe in the system', I believe that 'the system itself is at risk'; Gittelsohn 2016). With tension remaining at an all-time high, many savvy investors sense that sooner or later the spring will snap, and as confidence crumbles and the rulers run for the stock-market doors, a new MBM will get under way.

6.8 The temporal basis of confidence in obedience

In sum, we can say that the power and systemic fear indices tell us two opposite things about the capitalist outlook: the power index expresses the capitalists' *outer* confidence in obedience, while the systemic fear index indicates the *inner* capitalist apprehension that the mode of power might not last. Although they point in different directions, though, the two measures are intimately intertwined: the systemic fear index shows the temporal basis of capitalist confidence. When power is low, confidence is based on the future and relies on forward-looking capitalization. But as power rises, the basis of confidence shifts to the present – and this shift, which implies a growing reliance on current profit, spells the progressive breakdown of forward-looking capitalization and a deepening fear for the system's future.

7. Strategic sabotage

This discussion leads us to the third aspect of our triangular CasP model: the claim that capitalized power and systemic fear are both driven by strategic sabotage.[25]

7.1 Is the stock market pro-cyclical or countercyclical?

Needless to say, this claim is diametrically opposed to the mismatch thesis. The dominant view is that financial markets oscillate around their economic fundamentals and that the oscillation is *pro*-cyclical. The stronger the fundamentals, the greater the optimism and therefore the larger the overshooting – and conversely, the weaker the fundamentals, the deeper the pessimism and therefore the greater the undershooting. This is why market 'bubbles' are supposed to inflate when the economy is booming and deflate when it tanks.

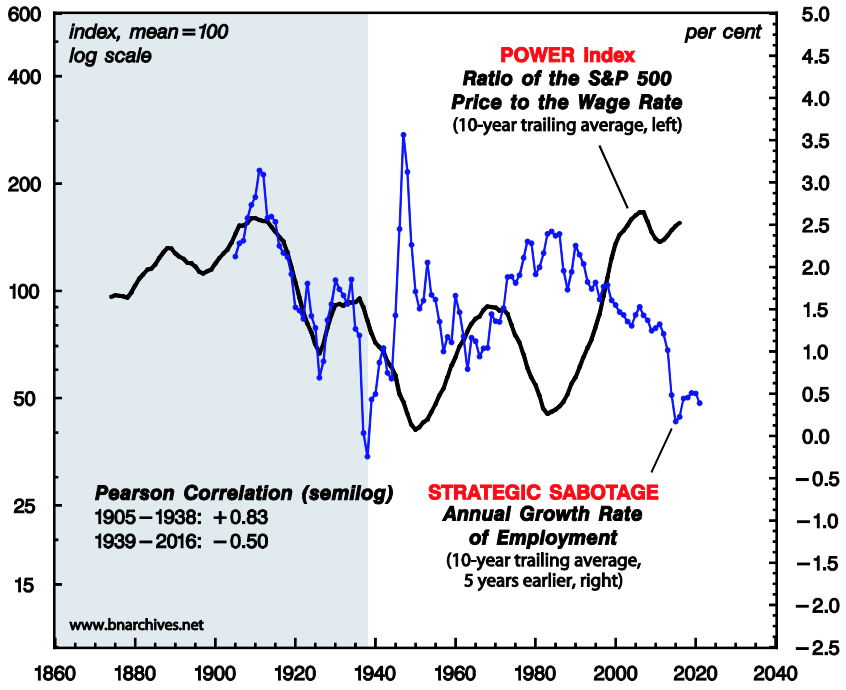
[25] The concept of strategic sabotage – the idea that rulers dominate society by strategically undermining, limiting and redirecting to their own ends the community's autonomy, creativity and productivity – was first articulated by Thorstein Veblen (Cf. 1904, 1923). This concept is central to our work on the capitalist mode of power (Nitzan and Bichler 2009a: Ch. 12) and has been examined, researched and extended in numerous CasP publications (for an outline of works on the subject, see Bichler and Nitzan 2015b).

Now, in our own work we have shown that, more often than not, the stock market is *not* pro-cyclical. As Figures 9 and 10 demonstrate, stock prices oscillate together with current profit only when the market is very 'expensive' (i.e., when capitalized power is extremely high). When valuations are moderately high, average, low or very low – which is the case most of the time – the movements of price and current profit are largely unrelated. Moreover, as we have demonstrated in our previous analyses of the mismatch thesis, the long-term growth rates of 'financial capital' (stocks and bonds) and the 'real capital stock' (measured in current dollars) correlate not positively but *negatively* (Bichler and Nitzan 2009, 2015a). What we would like to argue here is that this counter movement of the stock market and the so-called underlying economy reflects the *sabotage* underpinnings of accumulation.

7.2 Employment growth

Begin with Figure 11. The chart shows two series. The first is our power index, smoothed as a ten-year trailing average and plotted against the left log scale. The second is the annual growth rate of employment, which is also smoothed as a ten-year moving average, and which we plot against the right scale. Notice that the employment growth series is lagged five years. This lag means that, if there is a connection here, this series can be seen as a *leading indicator* or *predictor* of the power index.

Figure 11: **The power index and strategic sabotage**



NOTE: The S&P 500 price splices the following four sub-series: a combination of bank, insurance and railroad stock series weighed by Global Financial Data (1820-1870); the Cowles/Standard and Poor's Composite (1871-1925); the 90-stock Composite (1926-1956); and the S&P 500 (1957-present). The wage rate splices hourly data for manufacturing production workers till 1946 with hourly data for nonfarm business-sector workers from 1947 onward. Shaded area denotes positive correlation. The last data points are 2016 for the power index and 2021 for the lagged strategic sabotage.

SOURCE: The S&P 500 price is from Global Financial Data (GFD) till 1900 (series code: _SPXD) and from Global Insight (GI) from 1901 onward (series code: JS&PNS). The hourly wage rate splices the following series: *Historical Statistics of the United States*, Millennial Edition Online: hourly wages in manufacturing, all trades, 1865-1889 (series code: Ba4290), hourly earnings in manufacturing, all

industries, 1890-1913 (series code: Ba4299), weekly earnings of production workers in manufacturing, 1914-1918 (series code: Ba4362), hourly earnings of production workers in manufacturing, 1919-1938 (series code: Ba4361); Global Insight (GI): average hourly earnings of production workers in manufacturing, 1939-1946, (series code: AHPMFNS); Conference Board through GI: average hourly compensation of all employees in the nonfarm business sector (series code: JRWSSNFE). Employment till 1947 is from *Historical Statistics of the United States*, Millennial Edition Online (series code: CivilianLaborForce_Employed_Total_Ba471_Thousand) and from 1948 from U.S. Department of Labour through GI (annual average of monthly series, series code: ENS@US).

The reason we focus on employment growth is that it is a meaningful proxy from *both* the mismatch and power perspectives. From a mismatch viewpoint, employment growth is a *direct* measure of economic activity (unlike GDP growth, which is the *outcome* of that activity). From a power perspective, employment growth is an *inverse* proxy of strategic sabotage: it is one of the most crucial metrics of the wellbeing and sense of security of the underlying population, so to lower this proxy is to strategically sabotage most people.[26]

7.3 Switching the social current

Now, the thing that jumps out of the chart here is the remarkable 1939 reversal of the relationship between the two indices: until 1939, employment growth and the power index were positively correlated – and then, suddenly, as if someone switched the social current, the relationship turned negative.

How is such overnight reversal even possible? The answer to this question requires much more research than we can offer here, but let us outline our own initial thinking about it. Recall that the 1930s experienced an MBM, and that this MBM was ultimately resolved by creordering the entire mode of power. The chief hallmark of this creordering was the rise of the Keynesian welfare-warfare state. And one of the key pillars of Keynesianism was discretionary countercyclical economic policy – and specifically for our purpose here, countercyclical *monetary* policy. In our view, this dramatic shift toward discretionary countercyclical monetary policy might explain why the correlation between capitalized power and strategic sabotage suddenly inverted.

7.4 Employment growth and monetary policy

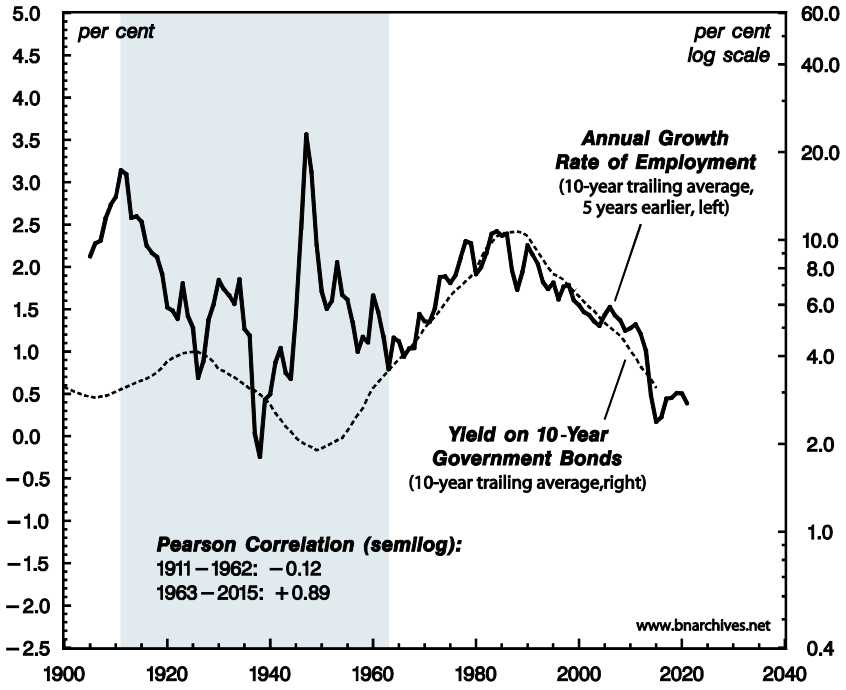
Let us examine this hypothesis a bit more closely. Until 1939, the power and employment growth indices were tightly and positively correlated (recall that employment growth is an inverse proxy of sabotage). The reason was twofold. First, capitalized power and systemic fear, although declining, were still relatively high, so changes in stock prices moved closely with changes in current earnings, and therefore with changes in employment (profit and employment levels are positively correlated). Second, since there was no countercyclical monetary policy, this positive correlation was largely undisturbed.

After 1939, though, capitalists started to expect government policy to mitigate the cycle, in part through countercyclical monetary policy. And since the cycle is driven by employment growth, they expected higher employment growth to bring about tighter money, higher interest

[26] The inverted growth rate of employment, although central, is not the only form of sabotage (see footnote 25). Many other processes serve to undermine the wellbeing and security of the underlying population – though incorporating them into our indicator might end up being a double-edge sword. While sharpening the analysis, they will likely make the indicator more complex, shorter in duration and difficult to apply in other countries.

rates and therefore lower capitalization. This expectation took a bit to materialize, but as Figure 12 shows, eventually it became self-fulfilling.

Figure 12: **Employment growth and the rate of interest**



NOTE: The last data points are 2015 for the yield on ten-year government bonds and 2021 for the lagged annual growth rate of employment.

SOURCE: Employment till 1947 is from *Historical Statistics of the United States*, Millennial Edition Online (series code: CivilianLaborForce_Employed_Total_Ba471_Thousand) and from 1948 from U.S. Department of Labour through GI (annual average of monthly series, series code: ENS@US). The yield on ten-year government bonds splices data from Global Financial Data (GFD) till 1953 and from the U.S. Federal Reserve Board through Global Insight (GI) from 1954 onward (series code: RMGB10Y@US).

The chart shows two series: the annual rate of growth of employment (lagged 5 years) plotted against the left scale and the yield on ten-year government bonds on the right. We can see that, until the middle of the twentieth century, there was little or no connection between the two series (Pearson coefficient of –0.12). But from the early 1960s onward, employment growth has been a nearly perfect five-year leading predictor for interest rates: the lower the rate of growth of employment (and therefore the greater the strategic sabotage), the lower the subsequent rate of interest (with a Pearson correlation of +0.89). And since lower interest rates boost capitalization, you can clearly see how strategic sabotage worked to fuel capitalized power.

7.5 From a political business cycle to a CasP policy cycle

During the early 1940s, Michal Kalecki wrote an important article in which he identified what he called the 'political business cycle' (Kalecki 1943). The business cycle, he argued, had been deeply politicized in more than one way, and was now increasingly driven by government policy. One of Kalecki's main points was that expansionary policy boosts

employment and wages as well as profits, and therefore serves the economic interests of *both* workers and capitalists. But the efficacy of such policy demonstrates that governments can replace capitalists in generating and maintaining prosperity, and that demonstration undermines the long-term *political* interest of capitalists in preserving their class rule. According to Kalecki, this contradiction between the economic and political interests of capitalists means that, when it comes to *maintaining* full employment, capitalists are likely to sacrifice their economic interest in higher profit for their political interest in continued dominance.

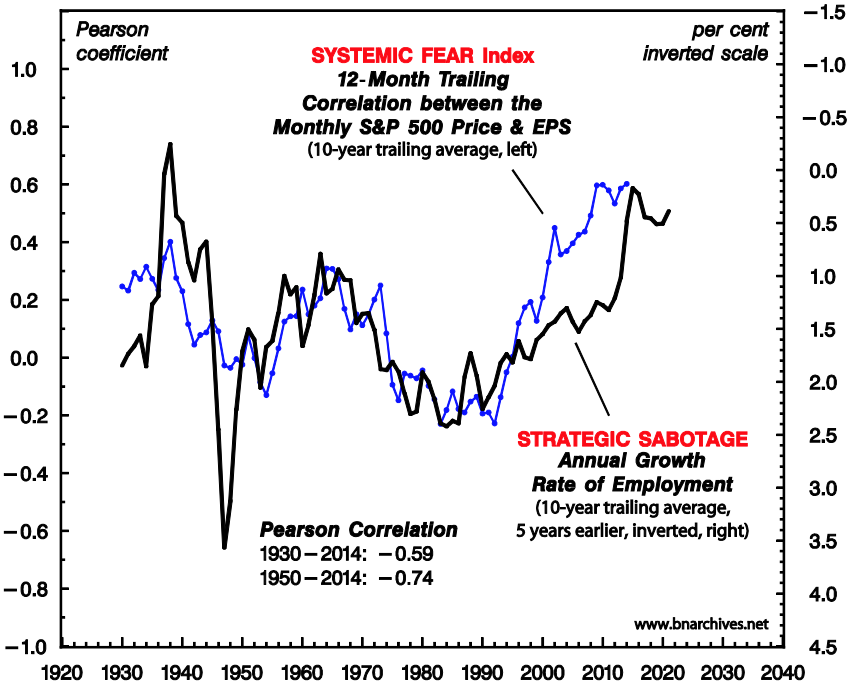
However – and here we come back to our own model – if we think of capital accumulation not as an economic activity, but as the capitalization of power, there is no sacrifice at all. The consequence of tighter policy may be lower employment growth and lower profit here and now – but, then, under normal circumstances, these short-term consequences have little or no bearing on forward-looking capitalization. At the same time, while the immediate *consequences* of the policy in terms of employment and profit have little bearing on capitalization, the *tools* of the policy – and particularly the rate of interest – have a huge impact. As Figures 11 and 12 indicate, intensified sabotage through lower employment growth has become a leading indicator for lower interest rates down the road, and lower interest rates boost capitalized power.

So what we end up with is not a narrow political business cycle à la Kalecki, but a broader *CasP policy cycle*. In this setup, the government is not some sort of exogenous distortion or an external alley but an integral component of the capitalization of power and the regulation of strategic sabotage. Moreover, there is no longer a conflict between the so-called political and economic interests of capitalists. Higher strategic sabotage keeps capitalists in the political driver's seat. And while this sabotage may undermine current profits, it allows lower interest rates, which are far more important for the capitalization of their forward-looking power.

Finally, in order to close our triangular model, Figure 13 shows the connection between strategic sabotage and systemic fear. This chart plots our systemic fear index on the left scale and the strategic sabotage index on the right scale. As before, both series are smoothed as ten-year trailing averages, and the strategic sabotage index is lagged five years. Note that the right scale is inverted to make the correlation easier to visualize.

Now, we already know that, since the 1940s, higher strategic sabotage has been associated, five years later, with higher capitalized power; but as noted, strategic sabotage also elicits resistance, and resistance raises the systemic fear of capitalists – exactly what Figure 13 demonstrates.

Figure 13: Strategic sabotage and systemic fear



NOTE: The systemic fear index represents annual averages of the monthly series shown in Figure 7 (see definition there). The S&P 500 price splices the following four sub-series: a combination of bank, insurance and railroad stock series weighed by Global Financial Data (1820-1870); the Cowles/Standard and Poor's Composite (1871-1925); the 90-stock Composite (1926-1956); and the S&P 500 (1957-present). The last data points are 2014 for the fear index and 2021 for the lagged strategic sabotage.

SOURCE: Employment till 1947 is from *Historical Statistics of the United States*, Millennial Edition Online (series code: CivilianLaborForce_Employed_Total_Ba471_Thousand) and from 1948 from U.S. Department of Labour through GI (annual average of monthly series, series code: ENS@US). Monthly S&P 500 price and EPS are from Robert J. Shiller's online data archives (http://www.econ.yale.edu/~shiller/data/ie_data.xls, accessed on April 20, 2016).

8. Conclusion

In summary, we started with the conventional creed of the mismatch thesis, which (1) assumes that equity investors constantly try to price stocks to match the underlying economic fundamentals; (2) posits that they rarely if ever succeed in doing so; and (3) shows that their presumed failure is a superb predictor of future returns.

We then proceeded to put this view back on its feet. We outlined a triangular CasP model of the stock market, showing that (1) mismatch valuation indices are in fact driven not by economic productivity and utility, but by capitalized power; (2) capitalized power is dialectically intertwined with systemic fear; and (3) both capitalized power and systemic fear are driven by strategic sabotage. An important corollary of this model is the notion of a CasP policy cycle – the idea that government policy, insofar as it caters to the imperative of capitalized power, favours low employment growth in order to enable low rates of interest.

And that observation leads us to the current historical moment. Over the past thirty years, U.S.-based capitalists (and others investing in U.S. equities) have managed to increase their capitalized power relative to the underlying population from record lows to record highs. In our

view, this increase has been driven by two related processes: (1) a redistribution of income from non-capitalists to capitalists, along with a growing conviction that the resulting inequality could be maintained and even augmented in the future; and (2) mounting strategic sabotage in the form of lower employment growth. In terms of Equation 2, the first process meant a higher hype coefficient H regarding the value of m. The second process has had a double impact: on the one hand, it assisted the first process by restricting wages and boosting profits, while, on the other, it enabled looser monetary policy and lower interest rates, thus helping to reduce the normal rate of return nrr (there has been no discernable decline in earnings volatility during this period, so it is hard to draw meaningful conclusions about risk perceptions).

These two processes were in turn underwritten by a major creordering of the underlying mode of power. Following the MBM of 1968-1981, capital has been progressively transnationalized, leading to the gradual disempowerment of the underlying domestic populations, the lowering of corporate and personal tax rates for high-net-worth individuals, the hijacking of macroeconomic policy for capitalized ends and the cajoling-forcing of pension funds and public assets into the stock market, among others consequences. But no spring can be pulled indefinitely. Conflict-driven redistribution and lower interest rates have pushed capitalized power toward its historical asymptote, and this approach means that the United States – and maybe the world as a whole – is now facing a historical crossroad.

Looking forward, we can see two possible trajectories. The less likely of the two is some version of Jack London's *The Iron Heel* (1907), in which the U.S. ruling class breaks through its historical asymptote by imposing a mode of power much harsher than the one prevailing over the past two centuries. To sustain this new mode of power, the rulers would have to further redistribute income in their favour, domestically and/or globally, leading to historically unprecedented levels of inequality. Moreover and crucially, they would have to cast this greater inequality as the 'new normal' (i.e., raise m in Equation 2) as well as persuade investors that this greater inequality is here to stay (so as to prevent hype H from collapsing). And while doing all of that, they would also need to keep interests rates and profit volatility low in order to prevent the discount rate from rising significantly – a tall order in a world marked by greater sabotage, intensified violence and therefore greater instability.

The other, and in our view more likely, possibility is that history will repeat itself, and that, sooner or later, the United States will experience another MBM. Now, if the past offers any guidance here, getting out of this MBM would require a major creordering of capitalism, domestically and globally, including the role of governments in the capitalization process. And here we come to our final and perhaps most important point.

Should it occur, this latter creordering is likely to unsettle the dominant dogma, and that unsettling might open a brief historical window for critical alternatives – new theories, novel experiments in public planning and radical proposals to undo capitalized power in favour of direct democracy and autonomy (for our own modest proposal on this issue, see Debailleul, Bichler, and Nitzan 2016). However, if these alternative theories, experiments and policies are to have any traction, they must *transcend the conventional fracturing of capitalism*. They have to overcome the outdated notions that capitalism is a mode of consumption and production counted in utils or socially necessary abstract labour time; that politics is distinct from and 'distorts/assists' the economy, as the case may be; and that finance is somehow a mismatched reflection of an underlying 'real' economy. To stick to these preconceptions is to

stay locked within the capitalist mindset. And if we remain locked in this mindset, we are bound to find ourselves, once the MBM has come and gone, in a new, capitalist-creordered version of the very same system.

References

[Most CasP-related publications are available for free from *The Bichler & Nitzan Archives* (http://bnarchives.net) and *Capital as Power* (http://capitalaspower.com).]

Aitmatov, Chingiz. 1983. *The Day Lasts More Than a Hundred Years*. Translated by John French, Foreword by Katerina Clark. Bloomington: Indiana University Press.

Anonymous. 1929. Fisher Sees Stocks Permanently High. Yale Economist Tells Purchasing Agents Increased Earnings Justify Rise. *New York Times*, October 16.

Bichler, Shimshon, and Jonathan Nitzan. 2008. Contours of Crisis: *Plus ça change, plus c'est pareil? Dollars & Sense*, December 29.

Bichler, Shimshon, and Jonathan Nitzan. 2009. Contours of Crisis II: Fiction and Reality. *Dollars & Sense*, April 28.

Bichler, Shimshon, and Jonathan Nitzan. 2010. Systemic Fear, Modern Finance and the Future of Capitalism. Monograph, Jerusalem and Montreal (July), pp. 1-42.

Bichler, Shimshon, and Jonathan Nitzan. 2012a. The Asymptotes of Power. *Real-World Economics Review* (60, June): 18-53.

Bichler, Shimshon, and Jonathan Nitzan. 2012b. Capital as Power: Toward a New Cosmology of Capitalism. *Real-World Economics Review* (61, September): 65-84.

Bichler, Shimshon, and Jonathan Nitzan. 2014. No Way Out: Crime, Punishment and the Limits to Power. *Crime, Law and Social Change.* 61 (3, April): 251-271.

Bichler, Shimshon, and Jonathan Nitzan. 2015a. Capital Accumulation: Fiction and Reality. *Real-World Economics Review* (72, September 30): 47-78.

Bichler, Shimshon, and Jonathan Nitzan. 2015b. The CasP Project: Past, Present, Future. *Working Papers on Capital as Power* (2015/04, December): 1-29.

Bichler, Shimshon, and Jonathan Nitzan. 2015c. *The Scientist and the Church*. World Economics Association Books: WEA.

Castoriadis, Cornelius. 1984. Value, Equality, Justice, Politics: From Marx to Aristotle and from Aristotle to Ourselves. Translated from the French by Kate Soper and Martin H. Ryle. In *Crossroads in the Labyrinth*. Brighton, Sussex: The Harvester Press Limited, pp. 260-339.

Castoriadis, Cornelius. 1991. Power, Politics, Autonomy. In *Philosophy, Politics, Autonomy. Essays in Political Philosophy*, edited by D. A. Curtis. New York and Oxford: Oxford University Press, pp. 143-174.

Cochrane, D. T. 2016. Who Has the Power? China's Commodity Use and the Canadian Mining Business. York University, Toronto. September 28-30. *Fourth Forum on Capital as Power: Broadening the Vista*.

Damodaran, Aswath. 2011. *The Little Book of Valuation. How to Value a Company, Pick a Stock, and Profit.* Hoboken, N.J.: John Wiley & Sons.

Damodaran, Aswath. 2012. *Investment Valuation. Tools and Techniques for Determining the Value of Any Asset*. 3rd ed, *Wiley Finance Series.* Hoboken, N.J.: Wiley.

Debailleul, Corentin, Shimshon Bichler, and Jonathan Nitzan. 2016. Theory and Praxis, Theory and Practice, Practical Theory. *Working Papers on Capital as Power* (2016/01, February): 1-13.

Fisher, Irving. 1892. [1965]. *Mathematical Investigations in the Theory of Value and Price*. Appreciation and Interest, 1896. New York: A.M. Kelley.

Fisher, Irving. 1896. What is Capital? *The Economic Journal* 6 (24, December): 509-534.

Fisher, Irving. 1907. *The Rate of Interest. Its Nature, Determination and Relation to Economic Phenomena*. New York: The Macmillan Company.

Gittelsohn, John 2016. Gross Trying to Short Credit to Reverse Decades of Instinct. *Bloomberg*, May 26.

Glassman, James K., and Kevin A. Hassett. 1999. *Dow 36,000*. 1st ed. New York: Times Business.

Harcourt, Geoffrey C. 1969. Some Cambridge Controversies in the Theory of Capital. *Journal of Economic Literature* 7 (2): 369-405.

Harcourt, Geoffrey C. 1972. *Some Cambridge Controversies in the Theory of Capital.* Cambridge: Cambridge University Press.

Hegel, Georg Wilhelm Friedrich. 1807. [1971]. *The Phenomenology of Mind.* Translated with an introduction and notes by J. B. Baillie. 2nd Revised ed. London and New York: George Allen & Unwin and Humanities Press.

Hobbes, Thomas. 1691. [2002]. *Leviathan*. Edited by Aloysius Martinich. Peterborough, Ont.: Broadview Press.

Hunt, E. K., and Mark Lautzenheiser. 2011. *History of Economic Thought. A Critical Perspective*. 3rd ed. Armonk, N.Y.: M.E. Sharpe.

Hussman, John 2000. *Hussman Investment Research and Insight* 1 (2, October 3).

Hussman, John 2015a. The "New Era" is an Old Story (and Introducing Market Cap/GVA). *Hussman Funds Weekly Market Comment*, May 18.

Hussman, John 2015b. Rarefied Air: Valuations and Subsequent Market Returns. *Hussman Funds Weekly Market Comment*, November 30.

Hussman, John 2016. Far Beyond Double. *Hussman Funds Weekly Market Comment*, October 31.

Kalecki, Michal. 1943. [1971]. Political Aspects of Full Employment. In *Selected Essays on the Dynamics of the Capitalist Economy. 1933-1970*. Cambridge: Cambridge University Press, pp. 138-145.

Kliman, Andrew, Shimshon Bichler, and Jonathan Nitzan. 2011. Systemic Crisis, Systemic Fear: An Exchange. Special Issue on 'Crisis'. *Journal of Critical Globalization Studies* (4, April): 61-118.

London, Jack. 1907. [1957]. *The Iron Heel.* New York: Hill and Wang.

Malik, Suhail. 2014. The Ontology of Finance: Price, Power, and the Arkhéderivative. In *Collapse Vol. VIII: Casino Real*, edited by R. Mackay. Falmouth, UK: Urbanomic, pp. 629-811.

Malik, Suhail, and Andrea Phillips. 2012. Tainted Love: Art's Ethos and Capitalization. In *Contemporary Art and its Commercial Markets. A Report on Current Conditins and Future Scenarios*, edited by M. Lind and O. Velthuis. Berlin: Sternberg Press, pp. 209-240.

Martin, Ulf. 2010. Rational Control and the Magma of Reality. Paper read at Eastern Economic Association Annual Conference, February 26-28, at Philadelphia.

McKinsey & Company, Richard Dobbs, Anu Madgavkar, James Manyika, Jonathan Woetzel, Jacques Bughin, Eric Labaye, Liebeth Huisman, and Pranav Kashyap. 2016. Poorer Than Their Parents? Flat or Falling Incomes in Advanced Economies. *McKinsey Global Institute* (July): 1-99.

McKinsey & Company, Tim Koller, Richard Dobbs, and Bill Huyett. 2011. *Value. The Four Cornerstones of Corporate Finance.* Hoboken, N.J.: John Wiley & Sons.

McKinsey & Company, Tim Koller, Marc H. Goedhart, and David Wessels. 2005. *Valuation. Measuring and Managing the Value of Companies*. 4h ed. Hoboken, New Jersey: Wiley & Sons.

Nitzan, Jonathan. 1989. Price and Quantity Measurements: Theoretical Biases in Empirical Procedures. Working Paper 14/1989, Department of Economics, McGill University, Montreal, pp. 1-24.

Nitzan, Jonathan. 1992. *Inflation as Restructuring. A Theoretical and Empirical Account of the U.S. Experience*. Unpublished PhD Dissertation, Department of Economics, McGill University.

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

Nitzan, Jonathan, and Shimshon Bichler. 2009b. Contours of Crisis III: Systemic Fear and Forward-Looking Finance. *Dollars & Sense*, June 12.

Nitzan, Jonathan, and Shimshon Bichler. 2014. Can Capitalists Afford Recovery? Three Views on Economic Policy in Times of Crisis. *Review of Capital as Power* 1 (1): 110-155.

Ostry, Jonathan D., Parkash Loungani, and Davide Furceri. 2016. Neoliberalism: Oversold? *Finance and Development* (June): 38-41.

Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Translated by Arthur Goldhammer. Cambridge, Mass and London, England: The Belknap Press of Harvard University Press.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2009. *This Time is Different. Eight Centuries of Financial Folly.* Princeton: Princeton University Press.

Shiller, Robert J. 2000. *Irrational Exuberance.* Princeton, NJ: Princeton University Press.

Singh, Simon. 2004. *Big Bang. The Most Important Scientific Discovery of All Time and Why You Need to Know About it.* London and New York: Fourth Estate.

Veblen, Thorstein. 1904. [1975]. *The Theory of Business Enterprise.* Clifton, New Jersey: Augustus M. Kelley, Reprints of Economics Classics.

Veblen, Thorstein. 1923. [1967]. *Absentee Ownership and Business Enterprise in Recent Times. The Case of America*. With an introduction by Robert Leckachman. Boston: Beacon Press.

Zweig, Jason. 2009. Be Inversely Emotional, Not Unemotional. *The Wall Street Journal*, May 26, pp. 28.

Author contact: Shimshon Bichler: tookie@barak.net.il; Jonathan Nitzan: nitzan@yorku.ca

SUGGESTED CITATION:

Shimshon Bichler and Jonathan Nitzan, "A CasP model of the stock market", *real-world economics review*, issue no. 77, 10 December 2016, pp. 118-154,
http://www.paecon.net/PAEReview/issue77/Bichler-Nitzan77.pdf

You may post and read comments on this paper at https://rwer.wordpress.com/comments-on-rwer-issue-no-77/

real-world economics review, issue no. 77
subscribe for free

A New "General Theory"?
A review of *Capitalism* by Anwar Shaikh

Bernard Guerrien [Université Paris 1, France]

> "The goal of this book is to develop a theoretical structure that is appropriate from the very start to the actual operations of existing developed capitalist countries. Its object of investigation is neither the perfect, nor the imperfect, but rather the real. For this reason, the theoretical arguments developed here, along with their main alternatives, are constantly confronted with empirical evidence" (p. 4).

The main claim made by the French students in the "Open letter from economics students" in 2001 was that they wanted "to escape from imaginary worlds". Anwar Shaikh's new book *Capitalism: Competition, Conflicts, Crises* meet this expectation. Shaikh takes reality as a starting point and then comes back to it in order to test the theory proposed – far from the imaginary world, not to say delirious world, of the neoclassical theory.

According to James K. Galbraith, "there hasn't been one like it for 150 years". Even if the reviews are usually laudatory, it would be hard not to agree with him. *Capitalism* is an outstanding book as it proposes a general theory – that encompasses both microeconomics and macroeconomics – that is systematically confronted with empirical evidence. Shaikh describes his theoretical framework as "classical" since it is largely inspired by the work of Smith, Ricardo, and above all, Marx. His theory is constantly compared to the other major economic theories – neoclassical, Keynesian and post-Keynesian. The importance given to the confrontation of each assumption with empirical evidence is a major feature of the book.

Moreover, none of the assumptions made are contradictory to common sense – as opposed to neoclassical models. This book could provide the basis for a radical reform of economics teaching – from undergraduate to postgraduate. This reform has been long-awaited by the students that are sick and tired of the current microeconomics and macroeconomics courses[1] – that is, by "stories" ("fables", "parables"…) embellished with mathematics. This reform is obviously unthinkable for the moment – especially because the ideological stakes are high. But one can always dream…

A dense book that is not easily accessible

Capitalism is a long and dense book – almost 1000 pages. The first chapter ("Introduction") gives a detailed summary, but a prior knowledge and a lot of dedication are required if one wishes to read it to the end. This review aims to provide an overview of *Capitalism* to show that this is a work of great interest and to make it more accessible. It does not have to be read from beginning to end. Some issues (theory of value, costs, consumption, multiplier, money…) can be picked and the related chapters can be read separately – with particular

[1] The content of these courses (without the mathematical elaboration) could be taught in history of economic thought

regard to their respective role in the theoretical framework given in this review. That is why the issues mentioned in it are followed by references – chapters, sections and pages in which a more in-depth analysis can be found. Because of the length of this review, some major topics like money, finance and international trade will not be addressed. They are examined in some (excellent) chapters that can be read independently. They do not play a critical role in Shaikh's theoretical framework except for the fact that the money supply is endogenous – which is now widely recognized, even by orthodox economists who still dither.

Shaikh belongs to the "classical tradition" that starts

> "from the bottom up, from the actual world that we observe around us, and then build abstractions from there" (Shaikh, 2016).

The notion of competition is one of these "abstractions" and Shaikh starts by reconsidering competition – which set him apart from the neoclassical and the post-Keynesians economists.

A sharp break with the models of perfect and imperfect competition

All economists admit that the model of perfect competition is not realistic. But only a few – even among heterodox economists – will concede that it is not relevant. They would rather introduce "imperfections" to make it more realistic. As Anwar Shaikh argues:

> "Orthodox economics operates within a hypothesized world of perfect competition in which perfect consumers and firms act to bring about supposedly optimal outcomes. The discrepancies between this model and the reality it claims to address are then attributed to particular imperfections in reality itself. Most heterodox economists seize on this fact and insist that the world is characterized by imperfect competition. But this only ties them to the notion of perfect competition, which remains as their point of departure and base of comparison."

In short:

> "There is no imperfection without perfection".

The myth – elusive but widely accepted by the heterodox economists – that perfect competition would have (more or less) prevailed in the early stages of capitalism is recalled in his introduction (chapter 1):

> "Heterodox economics generally accept the perfectionist vision as adequate to some earlier stage of capitalism but argue that imperfections rule the modern world. In either case, such approaches actually serve to protect and preserve the basic theoretical foundation, which remain the necessary point of departure and primary reference to an ever accreting list of real world deviations" (p. 4).

Therefore, Shaikh proposes to his "many Keynesian and post-Keynesian friends":

> "that we reject the claim that perfect competition was ever appropriate and refuse the notion that observed outcomes should be attributed to historically arisen imperfections" (p. 747).

His critique is also aimed at those who claim to be classical economists and yet accept this historical fiction:

> "Almost all modern classical economists treat the competitive firm in the same manner as neoclassical theory, as a price taker" (p. 18).

"Real competition"

Competition in capitalism has nothing to do with the neoclassical fairy tale, where nice "price taking" agents are kindly making offers and demands – acting as if they were alone in this world – and whose only motive is to make everyone happy by matching buyers with sellers.

Real competition in capitalism is "the war of all against all" (p. 14), that is, of labor against capital, capital against capital, labor against labor, seller against buyer, seller against seller and buyer against buyer.

Real competition generates a "turbulent" process whereby

> "the tactics and strategy of price-setting and cost-cutting firms in the face of shifting advantages and disadvantages created by their own interactions" (p. 363)

are coming into play. No one is safe from competition. Even the firms with substantial market shares are threatened by new entrants with lower costs and have to lower their prices or their margins.

Shaikh blames the post-Keynesians for viewing the world through the lens of "imperfection".

He contests their analysis, based on the neoclassical deduction that each firm that sets prices has a monopoly power – since expected demand has to be taken into account.[2] In "real competition", the capacity to set prices is distinct from the monopoly power: all firms set prices but this does not necessarily mean that they enjoy a monopoly.[3] This is a major distinction because if the majority of the firms behaved like monopolies (even "small" ones), profit rates would not tend to equalize – a key element in Shaikh's theory.[4]

Throughout the whole book, theoretical arguments are constantly confronted with empirical evidence. In the second section of the chapter 8 "On perfect and imperfect competition" dedicated to the "Empirical Evidence on Competition and Monopoly" (p. 367), he reviews

[2] As opposed to perfect competition in which the expected demand would be "flat", "horizontal" or "perfectly inelastic".
[3] Shaikh recognizes the existence of monopolies which are characterized by persistent excess (normal) profit rates
[4] Each firm defines a profit margin that depends on a variety of specific factors. There is no reason to believe that this profit margin should converge towards a common rate.

several studies on the measure of the profit rate and on its equalization tendency. He concludes that the empirical evidence is consistent with the theory of real competition.

A world where real competition prevails is naturally "turbulent". The war of all against all takes on many forms and the motives and behaviors of those who wage it are diverse. It is therefore misleading to represent this world with hypothetical individuals characterized by well-defined behaviors and motives – as the neoclassical economists do. The right way is to draw from reality that arises from the interaction of heterogeneous units and to try to detect "hidden structures", "lawful patterns" or "empirical laws". It is then possible to try to introduce "microfoundations" on the basis of a few constraints faced by the individuals. This approach is used in the chapter 3, "Microfoundations and Macro Patterns", that deserves attention since it is one of the distinctive features of Shaikh's book.

Micro foundations and emergent properties

The diversity of individual agent behaviors is a witnessed fact that has to be taken into account, not ignored: "Diversity must be embraced, not suppressed" (p. 9).

Even if they are complex, "whimsy" and subject to "occasional madness", consumer behaviors give rise to four "empirical laws": downward sloping demand curves, characteristic income elasticity of necessaries and luxuries, the nonlinearity of Engel's curves, and the near-linearity of aggregate consumption functions. Shaikh proves that these "laws" can be derived solely from two undemanding constraints: (1) a limited income and (2) a minimum level of consumption for necessary goods. He then proceeds to the simulation of four contrasting models of individual behavior and find that the four cases give essentially indistinguishable aggregate results.

He thus concludes that

> "aggregate outcomes are 'robustly indifferent' to microeconomic details" (pp. 96 to 101).

Hence,

> "the assumption of hyper-rationality (sic) is not useful because it systematically misrepresents the underlying motivations and is not necessary because we can derive observed patterns without it" (p. 8).

The same method is applied to production in chapter 14. The starting point is the observation that the division of the income generated by the production gives rise to a conflict between workers and their employers ("capital"). As with the consumption, Shaikh formulates two simple hypotheses that each firm must meet: the real wage must lie in between a lower limit and an upper limit (the former being "historically determined" while the latter corresponds to the productivity of labor – that is, the case where the workers get the whole surplus). If the conflict between capital and labor is (approximately) stable in firms – their distinctive features are relatively stable in time – then

> "disparate individual capital-labor struggles in a particular social climate lead to a particular ratio between the average wage and productivity" (p. 648).

This relation being

> "robustly insensitive to the micro details" (p. 648),

as with the consumption.

This relation between the average wage and productivity gives rise to a "classical curve" that plays a major role in Shaikh's theory – as we shall see later.

In the last chapter ("Summary and Conclusions"), Shaikh rightly argues that:

> "A central finding [of *Capitalism*] is that lawful patterns can emerge from the interactions of heterogeneous units (individuals and firms) operating under shifting strategies and conflicting expectations because outcomes are 'robustly indifferent' to microeconomic details" (p. 748).

Classical approach and operative principles

Observed "empirical laws" are not sufficient to build a theoretical framework: they can be used to verify the theory or they can be a part of it. Shaikh does not claim to start from scratch: he identifies himself with the classical tradition throughout *Capitalism* even if what is meant by classical is not specified. But we can progressively grasp what is implied, for example

> "The central goal of this book is to demonstrate that a great variety of phenomena can be explained by a very small of operative principles that make actual outcomes gravitate around their ever moving centres of gravity ...This is the system of *turbulent* regulation, whose characteristic expression takes the form of *pattern recurrence*" (p. 76).

Shaikh does not list the "operative principles" but three of them appear through the book*:* gain seeking behaviors, the "equalization principle" and the conflict between capital and labor for the division of value-added – to which we shall return.

The principles of equalization and of gain-seeking are jointly responsible for two emergent properties that play a critical role in the classical approach: price equalization and the profit rate equalization:

> "price and profit rate equalizations are *quintessential emergent properties*, unintended outcomes of constant jockeying for greater profits" (p. 14).

The "law of one price" thus appears as (approximately) resulting from a "turbulent" process and not as a premise – as in neoclassical theory.

All firms are involved in the equalization of the profit rate, including financial firms whose job is to make loans – as the banks. Bank profits are given by the interests earned on the loans that they make. Consequently, the interest rate gravitates around the general rate of profit –that is, the one that prevails in the economy. The interest rate is not determined by "supply and

demand" or by some kind of "liquidity preference". They can nonetheless be part of a "turbulent" dimension of the gravitation around the profit rate.[5]

For the firm,

> "the interest rate is the benchmark for the return on capital left passively in the bank rather than being actively invested in risky capitalist enterprise" (p. 443)

and this function does not depend on whether or not an individual firm borrows. Hence, as

> "Marx and Keynes emphasize, the investment is driven by the difference between the rate of profit and the rate of interest" (p. 23).

The tendency towards the equalization of the profit rates carries over to net profit rates insofar as we consider a common benchmark interest rate. Finance – where the risk factor is an additional consideration– is also subject to this equalization tendency. Therefore, the net profit rate is the driving force of investment and growth. In our "turbulent" world where different kinds of firms coexist, the net profit rate can vary from one company to another. As a matter of fact, the equalization process only concerns the most dynamic firms. That is why Shaikh introduces the notion of the "normal" level of a variable.

Real values and "normal" values

The distinction made between what happens at an abstract level and at a more concrete level leads to the two important notions of a normal profit rate and of a regulating profit rate.

> "While at the highest level of abstraction, competition appears to lead to a common technology and common price within an industry, that is to single point for each variable, at a more concrete level it can be shown to create and maintain a distribution for each variable" (p. 750).

That is because the different firms simultaneously use technologies that are relatively old or new. This affects their costs and in turn, their prices – the firms with newer technologies have lower costs and thus lower prices. For any given plant, new or old,

> "the normal capacity (potential) output is defined as the (normal) potential corresponding to the lowest average cost (cost being defined in the business sense)" (p. 32).

The "normal" rate of profit is therefore the profit rate obtained at normal capacity utilization. Given that

> "classical theory expects that new investment is embodied in the best generally reproducible plant and equipment" (p. 16),

[5] The interest rate is a source of considerable confusion for the neoclassical economists. In microeconomics, interest rates are real and specific to each good and each date – no money; in macroeconomics, it is confused with the profit rate (marginal productivity of capital). Keynesians and post-Keynesians have numerous theories of the interest rate.

it follows that there is also a normal rate of profit for the best generally available plant and equipment created by new investment in each industry. Shaikh calls these methods of production the "regulating" one because it is *their* normal profit rate which is subject to inter-industrial competition: an industry in which the normal profit rate of regulating capitals is above similar rates in other industries will attract more rapid investment flows and expand capacity relative to supply, thereby driving down prices and reducing the normal profit rate on regulating capitals. The opposite obtains when an industry's normal regulating rate of profit is lower than that in other industries. The end result is the turbulent equalization of regulating rates of profit. Hence for his theory of competition

> "the relevant measure [of the rate of profit] is the [normal] rate of return of new investment" (p. 16).

After explaining how to measure the rate of return on new investments, Shaikh examines the incremental rates of return on capital – which are a good approximation – across OECD industries. He concludes that:

> "in every case, average rates of profit tend to remain distinct while incremental rates of profit are strongly equalized" (p. 16),

which

> "provide considerable support for the classical hypothesis" (p. 16).

The equalization principle is thus operative – for profit rates and prices – provided that we focus on the appropriate profit rate. The profit rate of a firm largely depends on its costs. That is why Shaikh pays close attention to their measure – without dodging the issue of their relation with value and prices.

Value, prices and profit

The thorny question of the value and prices – of production and market prices – is not addressed until the chapter 9 ("Competition and Interindustrial Relative Prices") of *Capitalism.* After having reviewed the historical debate on this issue – especially focused on the transformation of value into production prices –, Shaikh analyses long time series on the US and shows that the measures of

> "market prices, direct prices (prices proportional to integrated labor times) and prices of production [...] give roughly the same results" (p. 21).[6]

The distances between the three measures are about 13 to 15% whereas the coefficients of determination are between 0,8 and 0,9. It can therefore safely be said that we can rely on the market prices to estimate the costs of firms. As with the "classics" and the "business practice", these costs are defined by

> "expenses on prime costs (material and wages) and fixed costs (amortization of fixed capital)" (p. 121).[7]

[6] In chapter 9, he points out that Sraffa says the very same thing (p. 440).

In chapter 4 on "Production and Costs", Shaikh revisits the neoclassical and post-Keynesian cost-curves – total cost, unit cost and marginal cost. He then develops his own theory that he relates to the classical tradition. As with post-Keynesians – nearly constant marginal cost, unit cost decreasing towards the marginal cost – special attention is given to the

> "length and intensity of the working day" (p. 121).

> And to the: "utilization of materials, fixed capital and labor" (p. 121).

Both have a significant impact on costs and hence on their measure.[8]

In the section VI "Empirical Evidence on Cost Curves" (p 160-164), Shaikh confronts his theory to several studies undertaken over the last decades to conclude that

> "the classical treatment of production is quite consistent with empirical evidence and that the theoretical cost curve derived on this base are similar to empirically observed curves and consistent with business experience" (p. 10).

In the chapter 7, Shaikh recalls that all the studies undertaken between the early 1930s – including those of the *Oxford Economists Research Group* and of Hall and Hitch - and the late 1990s – including those reported in the book *Asking about prices* of Blinder, Canetti, Lebow and Rudd – are at odds with the standard presentations. Unfortunately, the U-shaped average cost curves and the rising marginal cost still dominate textbooks[9].

A decisive parameter: the struggle between labor and capital

Labor is the only "good" that cannot be produced by capital:

> "Labor capacity is used by capital but it is not produced by capital" (p. 639).

As capitalists, workers have an autonomous power of decision. More specifically, the labor struggles play a significant role in the division of the value-added.

Unemployment is an inherent feature of competition in capitalism:

> "Competition creates a persistent pool (Reserve Army) of unemployment of labor" (p. 42).

---

[7] According to Shaikh, the neoclassical economists add to costs a "normal " gross margin. Indeed, the "authentic" neoclassical economists – those of the general equilibrium – don't for a an obvious reason: the profit rate is meaningless for them.

[8] Shaikh draws a distinction between the engineering capacity, which is the maximum production possible, and effective capacity – that is subject to regulatory and social constraints. As always, Shaikh focuses on the latter – that is, to those implemented by the businesses.

[9] Shaikh recalls how the "rebels" – Chamberlin, Joan Robinson, Sraffa, among others – who stood against the model of perfect competition in the 1930s have totally failed. After "neoclassical counterattack" – led by Stigler and Friedman –, perfect competition "ended up being elevated to new heights" (pp. 358-59).

Shaikh notes that not only the neoclassical but even the post-Keynesians, economist models conclude that workers are powerless in the determination of the wage-share:

> "a striking implication of both orthodox and heterodox approaches is that workers have nothing to say in their own standard of living. In the neoclassical approach, the real wage is determined by the full employment condition, in the post keynesian theory it is determined by productivity and the monopoly markup set by firms" (p. 42).

However, workers "have something to say" and their struggles must be taken into account in the analysis:

> "But once it is recognized that labor force and growth may themselves respond to accumulation through increases in labor force participation and/or immigration rates and through accelerated technical change, then there is full room for the effects of labor struggles for the real wage and wage share" (p. 42).

Shaikh's aim is

> "to develop a framework capable of accommodating the Keynesian and post-Keynesian understanding that accumulation is driven by profitability and that the aggregate demand has a central impact on output and unemployment" (p. 646)

along with the "classical recognition" that

> "labor struggles play a significant role in determining the real wage and that accumulation maintains a normal rate of capacity utilization alongside a persistent pool of unemployed labour" (p. 646).

Shaikh shows how a particular relation between the average wage and productivity can emerge from "disparate individual capital-labor struggles" in the firms (p. 648) by incorporating what he calls the "social-historical level of labor strength". Hence, he infers that *the rate of change of the wage share is a negative function of the unemployment rate*.

He calls this relation "classical curve" and adds it to the other "emergent properties".

The classical curve crosses the horizontal axis at a point corresponding to the "critical unemployment rate"

> "that will generally differ from the effective full employment rate" (p. 649).

In other words, critical unemployment – that for which the rate of change of the wage share is null – is, at least partly, involuntary (unlike the "natural" rate of unemployment of Friedman-Phelps and some "Keynesians"). The section VI of the chapter 14 is dedicated to "the relation of the classical wage curve to the Phillips curve". As usual, empirical evidence is considered.

Classical macro dynamics

The themes explored so far were mainly related to the "supply side" of the theory: production, costs, profit rate, labor struggles. The "demand side" is not addressed until the chapter 13 on "Classical Macro Dynamics" that starts with "a reconsideration of the theory of effective demand". Shaikh reconsiders the "micro foundations" of effective demand. He (partly) rejects the "Keynesian" idea that saving and investment are independent. This may be the case for the household savings, but not for the business savings:

> "The business component of the savings rate (i.e. the fraction of profit that goes into retained earnings) cannot be independent of the business investment rate, since both decisions are made by the same firm" (p. 604).

The assumption that

> "business savings are a fixed proportion of net income or profits implies that the business savings (retained earnings) are not linked to the needs of investment finance, which is contrary to business practice and empirical evidence" (p. 38).

Indeed, in the United States, where retained earnings are substantial,

> "the business savings rate closely tracks investment rate (to which it is roughly equal)" (p. 41).

Shaikh reminds us that a fully endogenous savings rate would imply a "zero multiplier" while the "full multiplier" holds if the savings rate is fixed. A "generalization" of the multiplier with a partially endogenous savings rate is presented in detail and with some mathematics in the Appendix 13 of *Capitalism.*

Another important implication for the theory of effective demand is that the successive waves of spending resulting in a multiplier effect are only possible if there is an extension of bank credit:

> "Less noted is the implication that each round in this process is generated by a fresh injection of purchasing power by bank credit because the excess of investment over saving in any particular round is assumed to be financed by bank credit." (p. 602)

A widening economic circuit necessarily implies a rise in business debt – which is the counterpart of the bank credit extension. Plus, the balance of power shifts in favor of labor – the reserve army falls. The combination of these two factors has an impact on the profitability of the firms and may lower future growth. This is probably the most controversial analysis in *Capitalism.*

A controversial statement

A demand stimulus will usually have two effects. On the one hand, it will boost production and employment. But on the other hand, it can lower the profitability of the firms and thus the growth rate. As Shaikh puts it:

> "Newly created purchasing power can pump up output and employment, just as Keynes argued, but [we will see] that this can lead to a reduction to the rate of growth. Then while short run output will be higher than it would otherwise have been, long-run output will be lower than it would otherwise have been" (p. 6).

He therefore considers that

> "the belief [of Keynesians/post-Keynesians] that persistent involuntary unemployment can be eliminated through appropriate fiscal and monetary policies" (p. 38)

is totally illusory. This may be acknowledged, even by persons who disagree with the previous point.

It may be impossible to permanently eliminate involuntary unemployment because it is inherent to the very working of capitalism. But it does not necessarily means that each stimulus will result in failure – by lowering the future rate of growth. We can consider that things are more complex – without sticking to the obvious fact that in cyclical economies, the booms are inevitably followed by busts. As a matter of fact, Shaikh is aware of it.

It is no coincidence that the last chapter's title is "Growth, Profitability and Crisis". Shaikh applies his theory to the post-war period in order to provide a "classical" reading of the economic crisis. It is unlikely that he reaches a consensus given the complex and sensitive nature of the issue. Nonetheless, he provides a solid basis for discussion, open to various developments – which is highly valuable.

(Minor) criticisms

Nobody's perfect. Some sentences in *Capitalism* are of doubtful validity – especially on central bank intervention or on the *Efficient Market Hypothesis*. But a closer look reveals that these sentences are ambiguously formulated, not erroneous.

There is one exception though: on some (important) points, Shaikh gives a distorted view – and sometimes an erroneous view – of the perfect competition model that lies at the heart of the neoclassical theory. He probably focuses too much on the textbook models.

1. Agents in the perfect competition model *are supposed to be "hyper-rational"* (a concept not defined by Shaikh) whereas it is quite the opposite that is true. Price taking agents, especially firms, are not "hyper-rational" but rather stupid or naive or

"myopic"[10] – the choice is yours. Indeed, neoclassical economists like to claim that the superiority of capitalism comes from the fact that the information given by the "price signals" emanating from "the market" is sufficient for an efficient allocation of resources.

2. "Perfect knowledge contradicts perfect competition" (p. 346). I disagree: *"perfect knowledge" (whatever that means) isn't an assumption of the perfect competitive model* – no matter what textbooks or Wikipedia claim. Thus, it cannot "contradict" perfect competition. Neoclassical economists only *assume* that firms *believe* that they will sell or buy anything at "given prices", and that their supplies and demands will not influence prices. Arrow and Debreu explain in their seminal article, *Existence of an Equilibrium for a Competitive Economy*, that they "*instruct* each production and consumption unit *to behave as if* the announcement of price *p* were the equilibrium value" (point 1.4.1, my italics)". These assumptions are "unrealistic", or ridiculous, but they prevent inconsistencies[11].

3. Like most heterodox economists, *Shaikh wastes time on minor criticisms and misses the major issue: if all agents are "price takers", who set prices*? The "mythical auctioneer" is mentioned – pp. 345-346, for example – but the issue is not sufficiently addressed. Shaikh even uses the image of the "large number" of consumers or "very small firms" (pages 17, 329, 341, 351, 357, 367, 368, 431) – a metaphor that neoclassical economists use systematically to avoid the question of price setting. By doing so, Shaikh implicitly validates the idea that the model of perfect competition is "unrealistic" whereas it is *irrelevant*. When he writes that "if the theory of perfect competition were empirically valid, all firms would have the same profit rate" (p. 272), *he implies that the model of perfect competition is relevant* – nobody would have the idea to confront an irrelevant model with empirical evidence[12].

4. Similarly, *Shaikh gives credibility to the New Classical models*. After explaining that they are of the Robinson Crusoe type (p. 565), he talks about "Pareto efficiency" (p. 581) – a notion that implies *at least two agents* –, "wages", "prices", "markets", "employment", "general equilibrium framework" (p. 582), "State", "economic policies" and "econometric tests" – as if it meant something in the world of a schizophrenic Robinson (Varian *dixit*). Even some prominent neoclassical economists (Solow, Arrow and Hahn, among others) observed that this is nonsense.[13] Shaikh lends even more credibility to these models by considering their empirical relevance. For example when he writes that "*Real Business Cycle* theory … is … far too weak to account for observed variations over the cycle" (p. 582), or that "in the end, RBC theory rests on weak empirical foundations" (p. 582).

All of this is unfortunate, but it does not have any impact on the overall quality of the book.

---

[10] Expression used by Shaikh in the first chapter : "Competition is taken to prevail only if there is a multitude of small (sic) price-taking firms each of which pursues its own myopic interest" (p. 17).
[11] An unrealistic theory can survive – there will always be some "epistemologist" to defend it with Friedman-type arguments. But an inconsistent theory cannot survive. That is why the neoclassical economists are past masters in the formulation of preposterous hypothesis.
[12] It should also be noted that the profit rate is not defined (the capital neither) in the model of perfect competition. Shaikh thus confuses micro and macro, endorsing the myth that models with an aggregate production function represent a "competitive" economy.
[13] Kirman's article "Whom and What Does the Representative Individual Represent ?" can be found in Shaikh's bibliography, but Shaikh doesn't refer to it when he examines "representative agent models".

Conclusion

When it comes to capitalism, two radically opposed views are possible. Capitalism can be seen as the war of all against all – and notably of workers against capitalists. It can also be seen as a "cooperation" between "factors" that contribute to the production – in that case, the coordination between agents is achieved by the "market". Shaikh embraces the former view while the latter is adopted by the neoclassical economists. Both visions are eventually tempered in order to match with the observed facts. Keynesians and post-Keynesians are in a relatively uncomfortable in-between position – given the exclusive nature of these polar views. Shaikh manages to build a "general theory" – in which issues are addressed both at a micro and macro level – that is based on few "operative principles" and "empirical laws" and that remains true to his initial vision. And that is a major achievement. Of course, we could argue that notions like "hidden structures", "turbulent regulation", "gravitation" or "emergence" are unclear and may reduce the scope of his theory and further developments. We could also express some doubts about the way in which data are processed to reach his conclusions. Finally we could criticize him because he does not come up with "solutions" to the major problems facing capitalist economies or because he only makes some general observations – derived from his theory – and no "predictions". But the same is even more true of other theories. Above all, no theory provides such a good foundation for thinking and discussion – that is, such a factual, rich and diverse one. *Capitalism* is a book that any self-respecting economist should, at least, have on his or her bookshelf.

Author contact: bguerrien@sfr.fr

SUGGESTED CITATION:
Bernard Guerrien, "A New "General Theory"? A review of *Capitalism* by Anwar Shaikh", *real-world economics review*, issue no. 77, 10 December 2016, pp. 155-167,
http://www.paecon.net/PAEReview/issue77/Guerrien77.pdf

ISSN 1755-9472

EDITOR: Edward Fullbrook,
ASSOCIATE EDITOR: Jamie Morgan

Articles, comments on and proposals for should be sent to the editor at
pae_news@btinternet.com
