sanity, humanity and science

real-world economics review

Formerly the *post-autistic economics review* ISSN 1755-9472

Issue no. 53, 26 June 2010 back issues at www.paecon.net

Subscribers: ***11, 457*** *from over 150 countries*

In this issue:

Real-World Economics Review Blog http://rwer.wordpress.com/

- Results of poll of German economists
- The Glass Wall
- Letter from 100 Italian Economists
- Out of equilbrium
- Bad economic policy still biggest threat to global economic recovery

real-world economics review on Twitter http://twitter.com/RealWorldEcon

May 6th --

Signals from a very brief but emblematic catastrophe on Wall Street

Paul A. David *Stanford University (USA), Ecole Polytechnique & Telecom ParisTech (France), UNU-MERIT (The Netherlands)*

Abstract

This essay begins by looking closely at the underlying structural causes of the discontinuity that appeared in the behavior of the U.S. stock market at 2:40pm in the afternoon of 6th May 2010, because the emblematic "catastrophic" aspect of the collapse of equity prices, and their subsequent equally abrupt rebound, renders these events potentially informative about things that can happen in a wider array of dynamical systems or processes – including those with consequences about which there is cause for serious concern. What transpired in those 7 minutes is viewed as being best understood as a hitherto unrecognized "emergent property" of structural conditions in the U.S. national stock market that all the actors in the story collectively had allowed to come into existence largely unremarked upon, through an historical process that was viewed generally as benign and therefore left to follow its own course of evolution unimpeded. The deeper significance of the events of May 6th lies in the attention it directs to the difference between a society being able to create and deploy technical "codes" enabling greatly enhanced connectivity for "exchange networks" - the condition of "hyper-connectivity" among an increasing number of its decentralized sub-systems, and a society that also provides timely mutually compatible institutional regulations and administrative rules for the coherent governance of computer-mediated transactions among "community-like" organizations of human agents. Regulating mechanisms operating to damp volatility and stabilize systems in which there is beneficial positive feedback are considered, as are a variety of circumstances in which their absence results in dysfunctional dynamic behavior. It is suggested that in view of the growing dependence of contemporary society upon on-line human-machine organizations for the performance of vital social and economic functions, continuing to focus resources and creative imagination upon accomplishing the former, while neglecting the latter form of "progress" is a recipe for embarking upon dangerous trajectories that will be characterized by rising systemic hazards of catastrophic events of the non-transient kind.

Keywords: "May 6th market break", price volatility, catastrophe theory, positive feedback, on-line transactions, computer-mediated communities, socio-technical system governance, symmetric regulation, regulatory by-pass, disruptive innovation

Contact Author: Paul A. David, Titulaire, Chaire Innovation et Régulation
Departement SES, Telcom ParisTech, 46 rue Barrault, Paris 75013, France
Email : pad@stanford.edu

May 6th --

Signals from a very brief but emblematic catastrophe on Wall Street

On the morning of May 6th the stock markets on the U.S. East Coast opened with the Dow Jones Industrial Average (DJIA) index at 10,867.80. Uncertainties about the prospective strength of recovery from the recession, heightened by worries about the repercussions of Greece's debt problems, and the impeding release of a U.S. jobs report, weighed upon investors during the ensuing hours of trading, resulting in a gradual downward drift of prices throughout the morning and early afternoon. By 2:00pm (EDT) the average of the 30 securities in the Dow Jones Industrials index was down 161 points – approximately 1.5 percent from its closing level of the previous day. Trading in a number of stocks on the NYSE had been slowed in the preceding half-hour by the exchange's automatic mechanism for suppressing volatility in individual stock price movements, and this condition spread so that during 2:00pm-2:30pm the number of stocks that were similarly affected had increased from 100 to 200.

There then ensued a sudden and extraordinary plunge in prices on the major national exchanges (see Figure 1, *ad fin*). At approximately 2:40pm prices of many stock began falling with extraordinary velocity, taking the Dow Jones Industrials index down to 10,445.84 by 2:42pm, a fall of 3.9 percent. In the next 2 minutes it fell another 573.2 points: the cumulative 998.5 point plunge in the DJIA represented a 9.16 percent fall and was the largest proportional intraday loss recorded by the index since the October 1987 "Market Break." Very much the same downward course was followed by the more broadly based Standard and Poor's 500 index (see Figure 1, *ad fin*). But the fright and pain experienced by investors watching an estimated trillion dollars of market value being thus wiped away during the market's "free-fall" turned out on this occasion to be both short-lived and quite limited. The Dow Jones industrials rebounded with still greater speed from the depths it had touched at 2:47pm, regaining 543 points within 90 seconds, so that the losses for the day had been cut to 4.26 percent by 3:00pm. When trading on the NYSE closed the DJIA stood at only a bit more than 3 percent lower than the level at which it had opened.[1]

The significance of this brief "Wall Street catastrophe" of 6 May does not derive from that episode's immediate or indirect financial repercussions. Rather, it should be seen in the *emblematic* aspects of the story that will be related here in response to the question "What caused this disconcerting free-fall of prices on the U.S. stock market?" To appreciate that quality it is helpful to start from the mathematical rather than the ordinary language meanings conveyed by the term *catastrophe.* For that present purpose it suffices to understand that the field of inquiry known as "catastrophe theory" (introduced by the French mathematician René

[1] The preceding account of the events of 6th May, along with the descriptive details of attendant events subsequently to be related, has been drawn largely from the material found in "Preliminary Findings Regarding the Market Events of May 6, 2010," *Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues*, May 18, 2010: 149 pp.[Available at: http://www.sec.gov/sec-cftc-prelimreport.pdf];and Mary L. Schapiro (Chairman of the U.S. Securities and Exchange Commission), "Examining the Causes and Lessons of the May 6th Market Plunge," Testimony Before the Subcommittee on Securities, Insurance, and Investment of the United States Senate Committee on Banking, Housing, and Urban Affairs, May 20, 2010. [Available at: http://www.sec.gov/news/testimony/2010/ts052010mls.htm.] See C. Nagi and M. Miller, "Electronic Trading to Blame for Plunge, NYSE Says," *Bloomberg*, May 6, 2010. [Available at: http://www.bloomberg.com/apps/news?pid=20670001&sid=aETiygQQ8Y3g], for the estimate of the market value erased during the plunge in stock prices.

Thom[2]) is a special branch of the study of *bifurcations*. The latter manifest themselves as sudden shifts, qualitative breaks or *discontinuities* in the behavior of dynamical systems. They are brought about by small, possibly imperceptible alterations in the particular circumstances of non-linear systems, which can cause stable equilibrium dynamics to disappear, or appear at some critical points – where not only the rate of change, but also the rate of change in the rate of change unexpectedly becomes undefined. Catastrophe theory, following Thom's original intentions, provides a general language for conceptualizing and modeling the dynamics of *forms*, rather than a "scientific theory" in the sense of Popper – because languages, like scientific paradigms are not propositions that can be submitted to testing and invalidation, but instead prove themselves to be useful, or not, in particular contexts and for certain purposes.[3]

A commonplace physical illustration of a "catastrophic event" – in this formal sense of the term -- may be experienced by letting your finger trace the surface of a draped fabric until it reaches a point where the surface (the "manifold" as mathematicians would speak of the shawl or cloak's three-dimensional surface) has folded under itself; there gravity will cause your finger's point of contact to drop precipitously from the surface along which it was travelling smoothly – to land upon the lower level of the drapery beyond the fold. That little passage is the "catastrophe." In the present context, what is especially relevant about this conceptualization of the "event" experienced by your finger is its generic nature: catastrophes thus conceived are not phenomena belonging to a category delimited by some size dimension of the system in which they occur, or according to the severity of their sequelae; nor are they to be uniquely associated with processes that that operate only in one or another range of temporal velocities (whether slow, or fast). Instead, the catastrophes to which this essay's title refers are *fractal*, possessing the property of self-similarity.

My premise in what follows is that to look closely at the underlying structural causes of the discontinuity that appeared in the behavior of the U.S. stock market at 2:40pm in the afternoon of 6th May will be worthwhile, because its emblematic "catastrophic" aspect renders it potentially informative about things that can happen in a wider array of dynamical systems or processes – including those with consequences about which there is cause for serious concern. It is fair, however, to give notice in advance that the quest for the cause of that mystifying stock price collapse may be unsatisfying, at least in one respect. Although it is likely that strenuous and persistent inquiries (such as those presently being undertaken by the staffs of the U.S. Securities and Exchange Commission, and the Commodity Futures Trading Commission) eventually will identify a proximate precipitating event in the observable chain of

---

[2] Renϑ Thom, *Stabilitϑ structurelle et morphogϑnΠse*. Reading, MA: W.A. Benjamin, 1972; Paris: InterIditions, 1977; *Structural Stability and Morphogenesis*, transl. D.H. Fowler. Reading, MA: W.A. Benjamin, 1975.

[3] See the illuminating non-mathematical discussion by David Aubin, "Forms of Explanations in the Catastrophe Theory of René Thom: Topology, Morphogenesis, and Structuralism," in *Growing Explanations: Historical Perspective on the Sciences of Complexity*, ed. M. N. Wise, Durham: Duke University Press, 2004, 95-130. Aubin (pp. 98-99) emphasizes that for Thom the catastrophe theory was a new method of modeling, aimed "to understand natural phenomena by approaching them directly, rather than relying on traditional reductionism. Its main concern was the creation and destruction of forms, but more precisely, as they arise at the mundane level of everyday life. Catastrophe theory posited the existence of a mathematically defined structure responsible for the stability of these forms, which he called the logos of the form. The models built with the help of catastrophe theory were inherently qualitative—not quantitative—and consequently were not suited for action or prediction, but rather aimed at describing, and intelligibly understanding natural phenomena."

reactions -- one of Aristotle's *causa per accidens,* no actor or actors will be found to have been culpable. What transpired in those 7 minutes will be better understood, instead, to have been a hitherto unrecognized "emergent property" of structural conditions in the U.S. national stock market that all the actors in the story collectively had allowed to come into existence largely unremarked upon, through an historical process that was viewed generally as benign and therefore best left to follow its own course of evolution unimpeded.

My broad contention, then, is that the genesis of this dramatic but transient episode is a reflection -- a small signal but a signal nonetheless – of the damaging potential created by the growing imbalance between two tendencies in modern society. On the one side we have a growing ability and drive to use technical code to build "hyper-connected" systems of dynamic interaction among the participants in novel transactional networks; on the other side, there is the indifferent degree of attention that too often is accorded to furnishing the resulting human and machine organizations with suitable institutionalized norms and regulatory codes. These would serve to damp destabilizing feedback effects that are likely to arise from mimetic behaviors and the relationships of complementary that have been enabled among the ensemble of actors (and *actants*).

If any lesson is to be drawn from the analytical retelling -- the *histoire raisoneé* of this "brief catastrophe on Wall Street," it points to the importance of the difference between computer-mediated "trading networks" or dedicated, special-purpose "exchange networks," on the one hand, and, on the other hand, actual "communities" whose members' transactions can be conducted with greater efficiency "on line." Digital technologies and the architecture of the Internet have greatly expanded possibilities of readily affording "connections" for message generation and transmission between "machines", and among peoples via machines. Unfortunately, however, the Web and its underlying layer of connected sub-networks forming the Internet do not automatically bring into existence functional human-machine organizations that possess the properties of "communities."

Unlike narrowly purposed exchange networks, "communities" possess developed structures of governance based upon social as well as technical mechanisms. The former require some substantial minimum degree of mutual recognition of the participants' common interest, enough to encourage reflexive identification of individual actors with the group. That, in turn, is likely to induce among the latter a measure of compliance with norms and mechanisms of social control sufficient to stabilize interactions among the individuals – keeping the ensemble operating within functional bounds. The complementary components of the resulting socio-technical structures, when properly designed and expeditiously implemented, can work together to regulate the dynamic interactions within the system supported by its underlying communication networks. In so doing it may enormously enhance the collective discretionary intelligence of the system as a whole. But, to achieve this requires giving due attention and resources not only to the technical engineering but to the development of appropriate social mechanisms of "community governance. It is perilous to proceed on the assumption that the essential minimum degree of regulatory compliance on the part of a network's participants will be spontaneously forthcoming, simply because so many are seen to be enthusiastically flocking to enjoy the benefits that hyper-connectivity has provided.

A search for the reasons why

What had happened? For a number of days following this "brief catastrophe," shaken traders, equity investors, and stock exchange executives remained uncertain and increasingly worried by the lack of any clear explanation of the episode, and the consequent possibility that the same phenomenon might soon recur; not to mention the even more worrying possibility that on the next occasion the "rebound" would not come so swiftly, and what would ensue was a descent to still lower and persisting depths in market valuations.

Reports of the inconclusive search for the reason why the market had collapsed filled the daily newspapers, the financial press and web services with initial rumors that pointed to suspected "trading irregularities," raising the specter of a concerted speculative attack on some stocks having been engineered among a cabal of hedge-funds. This soon was dismissed as a paranoid delusion that was wholly lacking in empirical foundations.[4] The field was thus cleared for the next hypothesis: the "fat finger" explanation. Evocatively titled, this posited a human error committed by a yet unidentified trader, who, it was said, had unintentionally submitted a computer instruction to sell $16 *billion* worth of stock futures, instead of a mere $16 million.[5] But inspection of trading records, and interviews with various hedge-fund managers conducted by the Securities and Exchange Commission and other agencies of the U.S. federal government in the days following, all failed to locate any localized action of the magnitude that would make it a plausible candidate for indictment as "the accidental culprit" that had triggered Thursday's shocking cascade of stock values. Attention then shifted to the possible consequences of the interactions between pre-programmed algorithms for high-frequency computer implemented trades in stocks and future equities capable of rapidly executing millions of transactions.[6] Without human interventions the connections among these could set in motion accelerating spirals of ascending or, as in the case it hand, descending prices.

This effort at explanation, however, was little more than an evocation by financial writers of the common association of the role that was attributed to computer program trading in the Black Monday stock market crash of October 19, 1987, which had seen the DJIA drop by 22 percent. Recollections of that not-so-distant historical experience's association with program trading were revived by the recent attention in the U.S. financial press to the new practices of high-frequency traders -- whose ever-faster on-line computers executed

[4] Inasmuch as some paranoids have real enemies, the utter lack of supporting evidence in this case was crucial in quickly quelling repetition of this and similar suspicions in the press and on Internet blogs.

[5] This particular rumor was making the rounds soon after the close of the NYSE on Thursday, May 6th. For example, the following passage appeared in a Associated Press news story by financial writer Mark Paradis, released with that date line and a 11:28 PM time stamp under the headline "Wall Street Roller-Coaster: Stocks fall 10 %": "No one was sure what happened, other than automated orders were activated by erroneous trades. One possibility being investigated was that a trader accidentally placed an order to sell $16 billion, instead of $16 million, worth of futures, and that was enough to trigger sell orders across the market." See, http://finance.yahoo.com/news/Wall-St-rollercoaster-Stocks-apf-892184148.html?x=0

[6] E.g., F. Norris, "Keeping up with the computers." *International Herald Tribune*, May 14, 2010. The theme of this genre of news story harked back to the diagnosis of the causes of the October 1987 "Market Break", which had followed the introduction of computer trading algorithms. High-frequently trading algorithms were cast in the role of the latest technical advance that harbored a renewal of the same dangers. But, as will be noted below, the implicit line of explanation omitted notice of the "spiral-halting" rules that had been put in place by the New York Stock Exchange and other exchanges in the wake soon after the 1987 "Market Break". On algorithmic trading, see the extensive entry and references in Wikipedia: http://en.wikipedia.org/wiki/Algorithmic_trading.

enormous volumes of programmed stock transactions at staggering speeds.[7] But the widely held view that implicated program trading as the main cause of the U.S. 1987 "market break" – even if the concurrent global stock price collapse evidently had other causes -- continued to be recalled in terms that inadequately acknowledged the consensus among financial experts that the heart of the problem at home lay not in the computers, but in the computer implementation of what then was a novel financial strategy known as *portfolio insurance*. Portfolio insurance today is a familiar method of hedging a portfolio of stocks against the risk of a market downturn by short-selling an index of stock futures; whereas in the years preceding Black Monday it was a novel financial technique that involved the use of a relatively new financial derivative -- bets on the future values of an index of average stock prices. Then, as now, this strategy was most frequently deployed by the managers of large institutional equity portfolios in conditions of market uncertainty, and rather less so when expectations are "bullish," because its use limits upside gains.

But what was revealed in the U.S. stock market on Black Monday was that program trading by portfolio insurers in conditions where equity holders' uncertainties suddenly were heightened holds a potential for catastrophic value destruction. Once the stock market began to drop, those who had long positions in stock futures would liquidate them, sending the prices of stock index futures downwards, while portfolio insurers who had sold the index futures short would move to sell their stocks, imparting another down-tick in the derivatives market (the stock future index), that would in turn trigger further stock sales, and so on, driving a continuing equity price cascade.[8] The painful discovery of the existence of that potential had prompted the introduction in 1988 of the device of "circuit breakers." These would halt trading on the NYSE when the fall in the average value of a market index reached a pre-specified measure, and this mechanism was extended by subsequent legislation and SEC regulations across the range of registered exchanges.

Preliminary inquiries -- carried out by the Securities Exchange Commission (SEC) and the Commodity Futures Trading Commission (CFTC) into the detailed dynamics of stock and future transactions on May 6th -- however, did not indicate that this particular positive feedback mechanism, involving the interaction between trades in stock index futures and sales of blocks of institutional portfolio-holdings involving stocks traded primarily on the major exchanges, was the source of the market break. The sharp rebound, and the observation that a disproportionately large number of the sales executed at bizarrely low prices had involved exchange traded funds (ETF) pointed in a different direction, because ETFs typically were

[7] Program trading had been introduced and quickly adopted in the U.S. during the mid-1980's, but still was not widely diffused abroad at the time of the global 1987 collapse of stock market values. Indeed, that movement had begun on the Hong Kong stock exchange (attaining a proportional amplitude twice that of the subsequent fall in the New York stock market), and it spread rapidly from there to other exchanges in the region where there was little program trading or none at all, as was the case in Australia. Evidently there were other causes at work in the larger episode than the one upon which American observers and analysts had particular reasons for focusing their attention. See U.S. Securities and Exchange Commission, *The October 1987 Market Break*. Washington D.C.:1988; also the Wikipedia entry at http://en.wikipedia.org/wiki/Black_Monday_%281987%29.

[8] Richard Sylla, a leading U.S. economic and financial historian, emphasizes that innovations with index futures and portfolio insurance had a critical role in the stock price cascade of October 19th: "I've seen accounts that maybe roughly half the trading on that day was by a small number of institutions with portfolio insurance. Big guys were dumping their stock. Also, the futures market in Chicago was even lower than the stock market, and people tried to arbitrage that. The proper strategy was to buy futures in Chicago and sell in the New York cash market. It made it hard -- the portfolio insurance people were also trying to sell their stock at the same time." See Annelena Lobb, "Looking Back at Black Monday: A Discussion with Richard Sylla," *The Wall Street Journal Online*. (October 15, 2007) Dow Jones & Company. http://online.wsj.com/article/SB119212671947456234.html?mod=US-Stocks.

traded on "thin", low liquidity, satellite markets of the major exchanges. This soon turned the attention of the SEC and the CFTC to examining the possibility that the cause of the market break might be found in unexpected and unusual losses of liquidity that had spread from stock to stock as a result of the interactions among the various "on-line' computer-mediated stock-trading *sub-systems* that form the national market system.[9]

In effect, the search for the reason why prices suddenly plunged and rebounded has brought into salience the little appreciated implications of a transformation that had occurred in the stock market's structure: a multiplicity of sophisticated electronic stock trading venues had come into existence in the preceding decade, forming a distributed system whose parts were not *fully integrated*, in the sense of operating under a symmetrically applied governance structure. The now-favored explanatory narrative of this episode, as I presently construe it, therefore runs along the following lines.

Although the ultimate origins of the precipitating price movement in the early afternoon still remain unidentified, it most probably was generated by computer-trading -- quite possibly of stock futures.[10] But the dynamic process that it precipitated, by triggering the New York Stock Exchange's mechanism for slowing the pace of trading in that venue, is of greater explanatory significance than the particular cause of the initiating shocks felt in some individual stock prices. The NYSE's trading system incorporates a "hybrid" mechanism that is intended to dampen volatility in a given stock by temporarily converting transactions in that title from an automated (electronic) market to a manual auction market. This action is triggered by the occurrence of proportional change in the price of the stock that exceeds a pre-specified size. The periods during which trading in the stock is thereby slowed are referred to as "liquidity refreshment points" (LRPs), and they can be as brief as one second, or even shorter in duration. The LRP's effect is to provide an opportunity for additional "liquidity" (in the form of buy, or sell orders) to enter the market and moderate the initiating movement of the stock's price. During these intervals, however, an electronic price quotation for the effected stock is displayed by the NYSE, even though it is not immediately accessible and may be by-passed by other trading venues and order routers – although such by-pass is not mandatory.

The NYSE's LRP mechanism, it should be noted, is quite distinct from the total, market wade "trading halts" that are triggered by large price movements in the major exchange averages, namely, those moving the Dow Jones industrials average by -10% or more. The mandated "circuit-breaker" action suspended trading not only on the NYSE but across all the registered stock exchanges and the equity futures markets. Once initiated, such circuit-breaking halts remain in effect for predetermined periods of time -- 30 minutes or longer, depending upon the magnitude of the Dow Jones industrials index's movement and the point in the trading day at which the triggering price change occurs. This system had been imposed by SEC and CFTC regulations in 1988, in the aftermath of the October 19the 1987 ("Black Monday") market break that saw the DJIA drop by 22 percent. It is significant that in

[9] See http://www.sec.gov/sec-cftc-prelimreport.pdf, esp. pp.25-33.

[10] The evidence for this, cited in the Senate sub-committee testimony of SEC Chairman Mary L. Shapiro, on 18th May, is that the precipitous decline in stocks followed very closely the fall in the E-Mini S&P 500 futures index (see Fig 1). The latter tracks the normal relationship between futures and stock prices for the S&P 500, a broader market index. See also the portion of the testimony of CFTC Chairman Gary Gensler, on 11th May, before the a sub-committee of the House of Representatives Committee on Financial Services [http://www.cftc.gov/PressRoom/Events/opaevent_gensler051110.html], dealing with trades in the E-Mini S&P futures contract around the critical period of the market's sell off.

approving the introduction of circuit-breakers in 1988, the SEC and CFTC explained that the rules were "an effort by the securities and futures markets to arrive at *a coordinated means to address potentially destabilizing market volatility*" such as recently had been experienced on "Black Monday," and not to prevent markets from reaching new and higher price levels.[11]

In 1998, following another stock market price collapse (that of October 27, 1997) the original rules regarding the magnitude of the absolute price declines required to trigger the circuit-breakers were reformulated in terms of (higher) proportional declines, upon which the duration the halt in trading were made conditional.[12] But, no halt occurred on May 6th, because the afternoon plunge in the DJIA did not reach the smallest (-10%) trigger point for a half-hour halt. Had the original trigger points not been modified by the SEC and CFTC in 1998, when those agencies accepted proposals by the NYSE and the Chicago Mercantile Exchange that specified a larger percentage decline (namely -10%), the May 6th plunge in equity and equity-futures prices would have been stopped by the suspension of trading. Pursuing that counterfactual proposition, however, would open the door to a speculative tale —rather than the one being related here.

When the NYSE slowed trading in the group of stocks whose prices had dropped, a number of the "market-makers" or "specialist" broker-dealers on that exchange would have found themselves committed to large-sized trades in those stocks for their institutional clients. Disinclined to hold onto this risk, they routed those orders to alternative, satellite electronic exchanges where trading still was ongoing at full speed. The latter venues, however, are not primary locations where under normal circumstances traders hold large order-books involving big blocks of institutional clients' stocks. Rather, they attract smaller retail trade orders that often can find better execution prices (and attention) there than they get on the NYSE or the NASDAQ. In comparison with those iconic stock exchanges these satellite markets are "thin" and at any given moment in the trading day may be quite "illiquid." Consequently, it is to be expected that the execution prices on large-size stock offerings received in those venues would be set at whatever low bids they will have happened to be holding. Were none to be found, the unfilled orders would instantly be forwarded on to other electronic satellites, where the process was likely to be repeated. In this way, further discontinuous drops in equity prices were generated, at a pace accelerated by computer-trading when waves of programmed "stop-loss" orders to sell were able to be executed. That would work to exacerbate the lag of trading on the NYSE behind the declining prices in those active but comparatively illiquid trading venues.

The slowed trading of affected stocks on the NYSE could not automatically have induced a spread of similarly stabilizing reactions across the other exchanges, simply because the "LRP" mechanism of reverting to manual auction trading was *sui generis* to the Wall Street exchange, and algorithms emulating the latter mechanism had not been implemented by satellite stock exchanges and the NASDAQ. In the absence of symmetric damping mechanisms throughout the national stock market system, certain features of the

[11] The quotation (with my emphasis) is from http://www.sec.gov/sec-cftc-prelimreport.pdf: p. C-3. See *ibid*., Appendix C for a chronicle of the evolving SEC and CFTC rules governing the use of these so-called "circuit-breakers" that during the period between 19th October and 30th December, 1998.

[12] According to the SEC – CFTC Staffs' Preliminary Report (*op.cit*., p.C-3), in 1988 these agencies "also believed that circuit breakers would help promote stability in the equity and equity-related markets by providing for increased information flows and enhanced opportunity to assess information during times of extreme market movements." This is in essence the rationale of the NYSE's LRP mechanism applied to movements in individual stock prices.

latter structure and trading practices among its participants that had hitherto appeared to be unproblematic, and under normal conditions were unproblematic if not beneficial, unexpectedly began to contribute further positive feedback augmenting the system's destabilizing dynamics.

One of these, noted in SEC Chairman Schapiro's testimony on May 18th before the Senate Sub-Committee,[13] was the interaction between the use of "market orders", "stop loss orders" and "stub quotes" in conjunction with the routing of unfilled orders from the NYSE to satellite exchanges, and transactions in thinly traded securities such as certain exchange traded funds and preferred stocks. A market order is an instruction to buy or sell a stock at the best available current price, and such orders do not require an execution at a specified price or price range. Generally, a market order is assured an execution, but what the execution price may be is not pre-specified. Therefore it is quite plausible to suppose that the use of "stop loss" orders on May 6th further amplified the downward impacts of market orders on prices, because stop loss" orders turn into market orders once the "stop price" of the order has been reached. In a normal market the use of a "stop loss" orders is a strategy that can protect an investor from incurring a major loss, but in the conditions that had developed on the afternoon of May 6th the market orders in conjunction with programmed "stop loss" orders had just the opposite effect, leading to executions at absurdly low prices.

This was only worsened by the practice of entering "stub quotes" in program trading algorithms – place-holder orders to buy a given stock at the lowest possible end of the price spectrum: a penny. Again, under normal conditions this practice is benign, since the probability of that price ever being reached for most stocks will be negligible. The practice of "stub quoting," however, has become much more extensive as a consequence of the spread of high-frequency trading: writing a stub order at 1 penny is particularly appreciated by high-frequency traders as a convenient way to be able to measure the round trip order latency period (the time lapse in getting an electronic order into the books) without tying up any material amount of the trading firm's capital. The entry of many lightly capitalized high frequency trading firms in recent years therefore tended to raise the likelihood that minimum stub quotes would be encountered when market orders were forwarded from the NYSE to thin satellite markets when LRP's had slowed transaction on the Wall Street exchange.

A second perverse effect emerged from the behavior of small high-frequency trading firms that had only recently entered the national stock market, and whose activities in normal times was seen as supplying liquidity to it. These firms' normal reactions to the movements of prices quoted for stocks on NYSE and its related electronic market NYSE-arca were altered by the abrupt elevation of the riskiness they now perceived in continuing their pre-programmed trading routines. Finding it increasingly difficult to make sense of the volatile streams of price data their computers were receiving from the major electronic exchanges, some small high frequency trading firms are reported to have worried that the NYSE-arca or the Nasdaq were experiencing technical malfunctions of their electronic trading systems. Were that to be the case in reality, transactions involving exchange originated orders eventually might be cancelled (or "broken") in compliance with the exchanges' policies.[14] That

[13] See http:www.sec.gov/news/testimony/2010/ts052010mls.htm, "Section 5.Other Factors."

[14] See e.g., J.Creswell, "In the blink of an eye, trades are made and questions are raised," *International Edition of the New York Times*, May 18, 2010. In the event, many trades on the NYSE and also some on the Nasdaq were cancelled after the event, not because there had been technical problems with the exchanges electronic systems, but exchange policy called for transactions executed at inexplicably deviant prices to be 'broken" to protect investor On "broken trades" and the detailed analysis of the

would leave the firm holding those positions when the prices of those securities eventually settled at much lower levels. These firms rarely if ever are "market specialists" and consequently are not under any obligation to continue trading and providing liquidity to the market, and could thus withdraw rather than hold the risk. But, because they were likely to have been (transiently) holding positions in millions of stocks, such decisions to close out their positions would reinforce the downward pressure on prices.[15]

As a result, the prices quoted for hitherto stable, "blue chip" corporate stocks suddenly plummeted --like that of Proctor and Gamble, which, quite inexplicably was more than halved within a few minutes -- sending understandably alarmed investors heading for exits. The development of this movement after 2:00PM saw a gradually growing number of stocks reaching "lows" in their inter-day values (see Figure 2 (*ad fin*), with the depths of those lows breaching the -20% range by 2:35PM. During the next 5 minutes price drops in the -40% to -100% range were registered for a handful of stocks. There was then a pause in the sale of stocks until 2:45 PM, after which timethe bottom abruptly fell out of the market and many, many stocks plunged to inter-day lows in the -40% to -100% range.

The cascade that pulled down equity values across the "Big Board" and the Nasdaq and most of other electronic exchanges was unsustainable, however, and once it slowed to a halt, the rapid rebound lifted the DJIA by 600 points, allowing it to close the day only 347 points (3.2 percent) below its opening level. Like a forest fire accelerating as it tore through new fuel, it burned itself out -- except that unlike a real fire that stopped when the fuel had been consumed and left a smoldering ruin, in this case much of the rapidly destroyed forest, magically, was quickly regenerated. Of course, there was nothing magical in what happened. Traders who had been short-selling equity futures now held positions that they needed to unwind, and their beginning to doing so initially checked the fall in prices of stock futures and brought buyers into the market for equities. After all, during half-hour of plunging stock prices nothing in the external environment of the securities market had been altered structurally that gave investors grounds to justify the bizarrely low prices at which trades in the stocks of many previously "solid" companies had been executed. As prices on external exchanges stopped falling, the number of LRPs on the NYSE dropped towards more normal levels and trading there began catching up with the movement of prices on the satellite exchanges. A positive feedback process emerged that was driving prices upwards, as market-makers came forward with orders to buy, which in pulling up prices there transmitted a mirroring impulse to the impulse to the equity-futures markets. With the external environment still essentially unaltered, the recovery could continue until stock portfolios had been rebuilt and the accompanying reappearance of portfolio insurance demands stabilized prices of equity-futures, stopping the interactive upward spiral.

---

distribution of securities in which these occurred in thinly-traded Exchange Traded Funds, see http://www.sec.gov/sec-cftc-prelimreport.pdf: pp.29-33, 42-44.

[15] On whether some high-frequency trading firms actually went off line after liquidating all their positions, see J. Creswell, *op cit*. Pulling the plug on their program trading activities reportedly was the reaction to bizarre price data from the major exchanges at Tradeworx, a small New Jersey-based trading firm whose founder was interviewed for this newspaper story.

A brief catastrophe, "long in the making"

From the foregoing story, it would appear that the brief catastrophe on Wall Street during the afternoon of May 6th actually had been a long time in the making. The structural conditions that were the "root cause" of the transient collapse -- and, more generally of the extreme volatility of stock price movements during this episode – are found in *the imperfect integration of the multiplicity of electronic trading venues* that collectively constitute what is now referred to (by the SEC) as the U.S. "national market system" (NMS). Many electronic exchanges have entered the market during the past decade's climate of deregulation and enthusiastic encouragement of innovation in financial instruments and institutions. Spectacular advances in the technologies of computer-mediated telecommunication were equally, if not more crucial in lowering the barriers to creating networked venues that could cater to the demand for "after-hours trading" in stocks and other securities, particularly the large demands that arise following the release of "news" (including corporate earnings reports, and other announcements) after the main markets have closed. This allowed the successive entry of small satellite exchanges that collectively grew to be major competitors of the large, long-established stock exchanges.

Once synonymous with the New York Stock Exchange, the Temple of Wall Street has thus become not simply "a virtual place", but the iconic representation of a "national stock market" within which its actual importance as a venue for stock transactions has been dramatically eroded. While the big broker-dealers that make the markets for the "large cap" stocks listed on NYSE's "Big Board" remain active at the historical site, the place has lost its former dominant position as the locus of publicly registered equity transactions. Having handled three-quarters of the volume of stock trades on U.S. exchanges only as recently as 2005, the NYSE's share (comprised mainly of NYSE and the electronic exchange NYSE-arca) had shrunk to approximately 35 percent of the U.S. registered exchanges in 2009, and presently stands at little more than 30 percent (See Figure 3, *ad fin*).

Those statistics leave out of the picture the so-called "dark pools" of private and unregistered equity trading venues, whose existence has only recently become visible to regulatory agencies like the Security and Exchange Commission, where belated concerns have been voiced that the large volumes of high-frequency trades being executed private for banks and hedge funds could destabilize the market. As a result of the loss of volume to its registered and unregistered electronic rivals, and the effect of industry-wide downward pressures on the margins of profits that the NYSE and other exchange operators are accustomed to extracting from the wedge between the prices facing the buyers and those facing the sellers, the profitability of the older established exchanges already had taken a beating well before the onset of the global financial crisis. The value of NYSE EURONext, the corporation that earlier in the decade successfully took the NYSE public (after its two centuries of ownership by the collectivity of private individuals who held seats on the Exchange) has been slashed by more than 75 percent during the past five years.[16]

[16] During the year ended December 31, 2009, EURONext operated under two segments: United States Operations and European Operations, providing various services in its United States and European markets. In the U.S. it operates the New York Stock Exchange (NYSE), NYSE Arca and a number of small satellites domestic exchanges including NYSE Alternext, NYSE Amex, NYSE Liffe US, LLC (NYSE Liffe US), NYSE Technologies, Inc (NYSE Technologies) and SmartPool. See Figure 3 *ad fin*, and http://topics.nytimes.com/top/news/business/companies/nyse-euronext/index.html]

But the “societal problem,” as distinguished from the business problems of competing stock exchanges, resides in the organizational fragmentation of the former stock market that has resulted in a largely unrecognized transformation of the financial landscape. The preceding references to the NMS’s “imperfect integration" alluded specifically to the fact that a significant number of the 50 electronically networked exchanges in the U.S. on which equities now are being traded did not follow the NYSE by putting in place mechanisms to damp large movements in stock prices. Nor were the supposed “liquidity-providing” firms that engaged in high-frequency trading obliged to stay in the market, rather than withdrawing when the riskiness of their positions suddenly appeared much greater than normally was perceived.

Consequently, what is of greater import in the current context than the NYSE’s shrinking share of the rising volume of market transactions, is how large a proportion of the potentially active traders in the market throughout the course of the ordinary trading day typically are attentive primarily to what is seen to be happening in that iconic venue rather than in the other exchanges. The relative number of regular traders, regardless of the size of the orders they place is likely to have remained much more concentrated on the NYSE (and perhaps a few other salient exchanges) than the volume of trades. That situation would have had a part among the conditions that gave rise of the dynamics that unfolded on May 6th, as it implies that a large number of potential traders were left lagging behind the ever-lower “bargain” prices being set on satellite exchanges – because their attention remained focused on the major, iconic NYSE. Under slowed condition of trading that obtained there as LRP’s affect a widening array of listed stocks, a comparatively small number of broker-dealers pushing large blocks of orders onto the thinner satellite venues contributed to individual stock prices breaking through previous “lows” and feeding back quotations to the NYSE that increased the volatility of price movements in the market as a whole.

The behavior of those traders who continued to fixate on the movements of shares on the NYSE and were slow to see what was happening in the less salient satellite exchange might have been simply a matter of previously formed habit, or a well calculated routine – since on-line monitoring has an opportunity cost because it competes for an individual’s attention implementing transactions. But even were one to suppose that the persistence of many traders in following the NYSE too closely was simply a path dependent routine, it is rather less reasonable to assume the same style of explanation for the persistence of so many of the satellite exchange operators in not following the NYSE’s decision to moderate the short-run volatility of stock-price movements.

Inaction in this one critical respect on the part of the electronic exchanges that had emerged as the NYSE rivals, patently was collectively destructive of the “public good” properties of liquidity and continuous adjustment that had formerly characterized the post-1987 U.S. stock market, and otherwise might have continued in a large network of electronically connected exchanges that were furnished symmetrically with volatility-moderating trading algorithms. In the aftermath of the surprising collapse on May 6th, it has struck many observers as quite – not to say inexcusably stupid -- that the newly registered electronic exchanges bothered to take the same precautions as those that had been in place on the NYSE, even as they were shrinking the latter’s share of national stock market transactions. Was it sheer indolence, or tunnel vision on the part of the new enterprises’ managers in adopting such collective measures that should bear the blame for the resulting "fragmentation” of the U.S. equity market's governance structure, or negligent hesitancy on

the part of the SEC and the CFTC to impose needed precautionary technology upon the growing number of registered exchanges?

To pose the issue in those terms, however, is to miss the point although high volatility in the prices and equities and equity futures has decidedly dysfunctional effects at the systemic, social level, greater price volatility hardly was perceived in those terms by all the new players that were entering the NMS. Rather, the business models of the operators of new electronic exchanges, like those of the small high-frequency trading firms aligned their respective private interests with the perpetuation of conditions of high price volatility that are characteristic of unregulated "thin markets." Larger price moments created greater opportunities for profitably exploiting technical advantages vis-à-vis other traders that high-speed programmed execution of large volume transactions in anticipation of the buying, or selling actions of others in the market. In that asynchronous form of temporal arbitrage, more abrupt price movements make for bigger profits. Not surprisingly, then, one hears from industry sources that "What all high-frequency traders love is volatility—lots of it."[17]

The answer to the question of why an asymmetric governance regime arose and persisted in the national stock market remains to be thoroughly researched, but surely it is pertinent to have posed it here, in the context of seeking to understand the micro-mechanisms implicated in the May 6th "market break." One striking implication of fragmentation of the governance structure of the U.S. stock market (a phenomenon that has not left the major stock markets of other countries, and their historic exchanges and bourses untouched) is that it rendered exchange stabilization measures of the sort that the NYSE had introduced not merely ineffectual, but likely to function *perversely* and thereby contribute to destabilizing stock price movements in the market as a whole.

To grasp the logic of that unintended but nonetheless bizarre outcome, consider what happens when trading a salient exchange like the NYSE is slowed during an episode of widening and more pronounced price fall. The net effect works to nullify potential bids from substantial "specialist" traders and market-makers in that venue. Those agents, having seen prices dropping to such aberrantly low levels that they would be moved to take a "long" position on their own account, find themselves delayed in executing the required purchase orders. In other words, with prices falling on satellite exchanges, the "speed bumps" that were slowing trading activity on the NYSE actually have the effect of inhibiting ("short-circuiting") the very reactions that otherwise should have worked to check the downward spiral of individual stocks' prices.

It is true that some investors and speculative traders have access to brokers who can deploy online tools in many alternative markets, allowing them to act rapidly upon expectations of a turn-around when prices have dropped to implausibly low levels. In principle then, their reactions should be able to work to halt a price collapse and thereby set the stage

[17] The quotation is from Creswell, *op.cit*. (2010), in a passage reporting an interview with a small high-frequency trading company, which continues: 'It was like shooting fish in the barrel in 2008,' said Majou Narang, the founder of Tradeworx. 'Any dummy who tried to do a high-frequency strategy back then could make money.'" When reading such comments it is relevant to note that beginning in the fall of 2007 and continuing throughout 2008, indicators of daily expected volatility in stock prices were elevated above the level that was characteristic throughout the preceding years, at least back as far as the beginning of 2005. That was the case, moreover, even before the November-May 2008 period of spectacularly high volatility. See the time-series of the daily values of the Chicago Board Options Exchange SPX Volatility Index ("VIX"), which uses options prices to construct a measure of the expected volatility of the S&P 500 stock price index, available at http://www.sec.gov/sec-cftc-prelimreport.pdf:p.12 (Figure 2).

for a rebound -- without having to engage the help of the mass of retail traders whose reactions via the NYSE also were delayed by the slower action of manual auctions. But, there is a hitch in that scenario: the greater likelihood is that many of the sophisticated participants in the market that are in a position to quickly avail themselves of the services of well-equipped brokers and high-speed trading programs would be focused on the very highly volatile movements occurring on the smaller and less liquid satellite exchanges. What they would have seen there during the critical "free-fall" phase would either induced them eventually to withdraw from trading entirely, as happened in some instances involving small firms, or left them more inclined to wait until the downward cascade of values began to slow -- sufficiently that the mass of other trader could be expected to soon take notice. Reaching that point would present sophisticated high-frequency traders with their opportunity to jump in ahead of the rest, and take their profits when the herd's arrival was driving the market's rebound. Yet, their waiting for the signal of that opportunity to become visible amidst the chaotic price movements emitted from the variety of illiquid satellite exchanges actually was allowing the free-fall to continue and thus postpone the onset of the market's eventual rebound.

Consequently, in the absence of a supervening regulatory authority willing to curtail the profitability of financial enterprises in the part of the market where the volume of stock transactions was growing most vigorously, there was little call for the introduction of stabilizing safe-guards. In any case, to provide cross-exchange stabilization of individual stock prices, the SEC would have had to break with precedent by extending its domain of its regulatory activity and require registered exchanges to install the algorithms needed to emulate the NYSE LRP system to provide cross-exchange stabilization of individual stock prices.[18] Such a regulatory initiative, moreover, would have had to be pursued in the face of presumptive claims that high-frequency traders were on balance net suppliers of liquidity to the market – a contention that is rather easier support when their actions' effects are being considered against a background of "normal" market conditions.[19] Nevertheless, the passivity of the SEC in this matter is made all-the-more striking by its contrast with the actions taken to protect against extreme swings of contract prices in the market for stock index futures. There, an array of automatic price-stabilizing control mechanisms that have been introduced by the Chicago Mercantile Exchange (CME) and the International Commodity Exchange (ICE), the two major exchanges on which the E-Mini S&P 500 futures and other stock index derivative are traded.[20]

---

[18] Such a step would have gone beyond the precedent set by the SEC in mandating the successive introduction (beginning in March 2007) of "penny-pricing" of exchange-traded stock options – starting with a selected group of electronic equity-options exchanges and eventually extending to all of them. The SEC justified imposing penny-pricing as beneficial for investors, and had to contend with the perception by some brokers in the exchange-traded stock options market that it would so narrow their margins and could only be in their interests if the volume of transactions rose sufficiently on all the exchanges. For the background of this innovation-diffusing action by the SEC. and its consequences for the brokers and the exchanges, see Elizabeth Stone, "The SEC and Electronic Options Trading, 1998-2008." Economics Department Seminar Paper, Stanford University, March 24, 2009.

[19] In the wake of the revelation of the consequences of not having cross-market volatility suppressing safe-guards in place, it might be reasonable to anticipate there now would be a voluntary movement toward symmetrically equipping all the registered exchanges with LRP-emulating programs and common regulations for their activation. On the other hand, the positive private business reasons for preserving volatility that have been identified could very well re-surface, engendering recalcitrance if not active resistance on the part of the new electronic exchanges and high-frequency traders – thwarting remedial regulatory reform of that kind.

[20] On both the CME and ICE, automatic rejection of orders falling outside of the pre-specified "reasonable" price ranges, and maximum size-limits on orders, as well as required lower limits for the execution of stop orders," all serve as protection against errors in entry orders, and endogenously generated extreme swings in prices. In addition, the CME Globex's "Stop Spike Functionality" automatically pauses trading for pre-specified intervals in order to protect against cascading "stop loss"

Reading the story of May 6th as a parable for our times

The deeper significance of May 6th's "brief catastrophe on Wall Street" – which is to say, its emblematic value – consists in the attention it directs to the difference between a society being able to create and deploy technical "codes" enabling greatly enhanced connectivity, the condition of "hyper-connectivity" among an increasing number of its decentralized sub-systems, and a society that also provides timely mutually compatible institutional regulations and administrative rules for the coherent governance of transactions among the human and machine agents throughout the multiplicity of the sub-systems upon which it vitally depends. Being far more able to accomplish the former than the latter form of "progress" is a recipe for a society moving onto dangerous trajectories characterized by rising systemic hazards of catastrophic events of the non-transient kind.[21]

The analogy been "digital code" and legal regulations (or more generally institutionalized social procedures and social "norms") governing human transactions, and the potential problems of permitting the former to be substituted for the latter on the Internet was influentially articulated more than a decade by Lawrence Lessig. The focus of the concerns originally animating Lessig's 1999 book, *Code* (now available in its 2nd Edition, 2009), was the transfer of rule-setting from the open political area, where a range of societal values might be reflected, to the province of a self-appointed technical cadre. The technocracy might not be so different from the politician in serving their own personal interests or conveniences or those of others who were animated by private commercial goals, but, more troublingly, were able to do so in a far less transparent manner without being held publicly responsible for the consequences. In a similar fashion, the continuing advancements in digital information technology not only permit but make it commercially attractive to go only pursuing greater degrees of technical connectivity, thereby facilitating interactions among human and machine subsystems that can form technically interoperable "exchange systems" without attaining socio-technical compatibility." This state of affairs implies that a rising portion of the digital information transactions upon which modern societies are growing to depend are subject to dynamic processes with non-transparent properties, yet remain ungoverned by social and technical norms that are regularly adjusted to maintain predictable functional stability compatible with the sustainability of the whole system.

The nub of the problem here is that IT innovation is being driven in part by attractive private incentives for entrepreneurial agents seeking ways to escape the constraints that

---

orders. Further details of these practices are given in CFTC Chairman Gensler's testimony before the House sub-committee [http://www.cftc.gov/PressRoom/Events/opaevent_gensler051110.html].

[21] As tempting as it is at this point to draw the connection between the present discussion of the case for institutionalized as well as purely technological measures to control self-reinforcing, positive-feedback driven market dynamics, and the grave challenges posed by similar processes at work in global warming, that digression to a far more complicated to subject must be resisted firmly. I have not been able, however, to exercise the same measure of self-restraint on other, previous occasions: see P. A. David, "From the Economics of QWERTY to the Millennium Bug, and Beyond: *Just-in-Time* for Y2K, and Next…For Irreversible Global Warming and Environmental Catastrophe?," presented at the Conference on *Historical Approaches in Economics*, held in Honor of Gavin Wright. Stanford Institute for Economic Policy Research, Stanford University, 27-28 September 2008; K.J. Arrow, L. Cohen, P.A. David, et al. "A Statement on the Appropriate Role for Research and Development in Climate Policy," *The Economists' Voice*, 6 (1):Article 6, March 2008 [DOI: 10.2202/1553-3832.1518. Available at: http://www.bepress.com/ev/vol6/iss1/art6.]; P.A. David, C. Huang, L. Soete, and A. van Zon, "Towards a Global Science and Technology Policy Agenda for Sustainable Development," *UNU Policy Brief No.2*, November 2009. [Available at: http://www.merit.unu.edu/publications/briefs.php].

socio-political regulations impose upon their profit-seeking activities -- when the latter are construed with reference to the exploitation of pre-existing structures of technological knowledge. One recourse open to those in that position is to seize opportunities to enter expanded "exchange systems" some aspects of which naturally, or by design extend beyond the pre-defined scope of existing regulatory controls – thereby altering the rules of the game to which they will be subject. Novel telecommunications technologies that create unfamiliar transactional networks, and especially those that can be connected with other subsystems to create new "hybrids" – by means of technical interface devises, or communication protocols, and intermediating agents – frequently have been especially doubly attractive, in offering both superior technical capacities and the prospect of being able to exploit these when liberated from the regulatory restraints that have come to be placed on long-established modes of production and forms of business organization. Indeed, it is in part because the participants entering a novel pure "exchange system" generally are not obliged to submit to a universal set of regulations that "standardize" their behavioral norms and technical practices, that the formation of such systems – which also offer opportunities for specialization and trade, agglomeration externalities, and arbitrage – have proved so attractive in the history of human societies.

This being the case, in the absence of external interventions by super-ordinate regulatory authorities, or the exercise of greater bargaining power by the sub-system (or network) with which the member of other subsystems (networks) seek to establish communications, the components forming exchange systems are likely to remain substantially differentiated in their internal workings. When interactions among them are dominantly complementary, so that the effect of exchanges is beneficial and consequently self-reinforcing for each of the subsystems, it is likely that some if not all of the subsystems will be exposed to what for them will be unprecedentedly strong positive feedback, self-reinforcing impulses that overwhelm whatever damping regulatory constraints or social norms had formerly been in within each of them. As a result, those enjoying substantial net advantages of having gained the requisite degree of "network connectivity" sufficient to support "exchange," or those entering such a system with relatively less effective stabilizing norms, are most likely to become the drivers of positive-feedback effects. That will only render the behavior of the emerging exchange system "predictably more uncertain" – in the sense of being less stable, or "excessively volatile" than that experienced formerly by each of its constituent elements. What is created is a world where there is less "safety in numbers," because heterogeneity among the connected actors tends to rise with their numbers, so that even though diversity may promote creativity within the system, unregulated innovations bring growing and non-transparent systemic risks.

It is useful, then, to see in the explanation of the recent minor episode on Wall Street a little parable, and perhaps also an instructive signal – if we are attuned to receive it – of that much bigger story. The moral I would have us draw from that tale is that the sheer pace of advance in digital information technologies has posed a species of governance problem to which polities and will have to learn to cope, or be more and more frequently beset by catastrophic collapses in the functions of the socio-technical information systems upon which the world's population has become increasing dependent.

The parable taken farther afield, across the spectrum of unwelcome 'emergences'

The problems of the unexpected emergence of volatile market behaviors with which the foregoing pages have been occupied, should by now be understood to have the potential to appear where-ever hyper-connectivity -- in the sense of dense highly efficient communications that facilitate exchange transactions among agents (humans and/or machines) -- allows the formation of a "system" to be technically enabled. But, destructive *sequelae* other than those that have been the center of the discussion thus far can be seen to issue from structural circumstances that produce "run-a-way" dynamical systems. Although some among these manifest themselves at the microcosmic level of personal interaction processes, the features they share with the sources of macroeconomic and financial instability are striking. A good parable is a highly elastic thing, in that it may be usefully stretched to fit the scale of large events that play themselves out over prolonged time-spans, or, as in the fleeting stock market break just examined, are compressed within the span of an hour. To put this point somewhat differently, one may say that the structural conditions permitting the self-referential inflation of "bubbles" until they no longer can be deflated and therefore must burst, are fractal. Accordingly they can generate analogous dynamical behaviors in a wide variety of circumstances, and at both the high-frequency and low-frequency ends of the narrative spectrum.

To explore this a bit more specifically, one may briefly turn to the sphere of quotidian and short-lived human communications facilitated by advanced (digital) telecommunications, and consider the congruence between the conditions seen to underlie the behavior of prices in the U.S. stock market on May 6th, and the phenomenon known in email forums on the Internet as "flaming." An episode of "flaming" refers to a behavioral break in the course of a discussion or negotiation conducted by email – via a "forum', subscription list, wiki, or game servers such as XBoxLive. Typically such incidents are marked by one of the participants descending precipitously into hostile, abusive commentary directed at another, or several others. These outbursts sometimes are reciprocated rather than remaining one-sided, and they are reported now also as occurring in the course of instant messaging exchanges and video-sharing websites. Going still farther afield, one might include as part of the same broader phenomenon the recent reemergence on the Web of an nastier, digitally implemented version of what in a former time were called "slam books": that being a reference to the medium in which teen-age school children (girls mostly) inscribed and circulated anonymous remarks about each others' perceived personal defects.

In the latter cases of quite corrosive exchanges, which proceed rapidly from disparagement to highly abusive comment, total anonymity of the parties seems the essential condition for sudden departures from civil intercourse between class-mates and ostensible friends; whereas, in other circumstances it is the parties' quasi-isolation from each other that seems to remove inhibitions on verbal aggression; either that, or the reduced likelihood of their unwelcome messages electing prompt negative feedbacks in the form of physical retaliation by the recipients.[22] Isolation, of course, is simply the extreme state of low

[22] See, e.g., N. A. Johnson, "Anger and flaming in computer-mediated negotiations among strangers," *Decision Support Systems* 46, (2009): 660-672, in which the conditions promoting this phenomenon are distinguished from those of computer communication *per se*, and contrasted with those in which computer-mediation reinforces a sense of community among the participants. "The literature suggests that, compared to face-to-face, the increased incidence of flaming when using computer-mediated communication is due to reductions in the transfer of social cues, which decrease individuals' concern for social evaluation and fear of social sanctions or reprisals. When social identity and ingroup status are

dimensionality channels of communication with others -- such as speaking to a person without being able to see their facial expressions or "body language," exchanging scripts without hearing the other party's intonations or involuntary utterances that usually are sufficient to cue adjustments in the tenor of human conversations; or, again, "texting" under symbolic restrictions that tightly limit the conveying of qualifications and nuanced expressions. In the latter case we have a tidy instance of an inexpensively effected "connectivity" producing exchange networks whose technically efficient performance is achieved by attenuating the affordances that contribute to the experience of "community."

Fortunately, these may turn out to be largely transient problems of the early Internet era, inasmuch as the availability of greater bandwidth at reduced costs could mitigate the "distancing" effect of text communications conducted with minimal grammar and vocabulary. But comparatively narrow bandwidth of greatly cheapened new communication media affording connectivity among geographically dispersed and anonymous parties do no not promise those salutary, mitigating effects and they bring reduced expectations of retaliation for disseminating hurtful messages. In this domain, however, unlike those involving pure machine organizations, it is *the anticipation* of negative feedback that serves as an important regulator, or volatility-moderating mechanism to limit damaging interactions among incompletely communicative parties.

More generally, even though humans are socialized from infancy to recognize precursory signals of adverse reaction by other to their behavior (such as shifts in facial express, oral tone, etc.), their interactions undergo perceptible transformations when they are insulated from the full range of their actions' adverse effects upon one another – either by spatial separation that limits the possibilities of physical retaliation, or by low-dimensionality and high latency in the channels afforded form message exchanges. Misunderstood messages that could be construed as hurtful are more likely to elicit responses in kind, rather than requests for clarification. There is thus a tendency to allow ourselves more scope to provoke those with whom we find ourselves exchanging messages, and in turn to reciprocate others' provocations. Liberated from the automatic triggering of "dampers," the back-and-forth process of verbal or symbolic messaging thus can, and often does proceed with escalating ferocity until the parties are separated.

There is a striking resemblance between this aspect of human social interactions -- when the actors can be connected to exchange messages without having to fully internalize the effects of those messages on the correspondent parties, and micro-biologists' observations on the evolution of virulence in microbial pathogens.[23] Comparison of viral, bacterial and protozoal agents of human diseases has shown that vector-borne pathogens have greater per-infection lethality than pathogens directly transmitted by contact with hosts. This is found to account for the severity of malaria, yellow fever, dengue, sleeping sickness and other insect-borne diseases being much greater that most of the respiratory-tract infections in humans. Pathogenic microbes (such as *Mycobacterium tuberculosis* – originally referred to as the "the "tubercle bacillus") that are transmitted directly by contacts between mobile human hosts have evolved in the direction of reduced virulence. Consistent with this

---

salient, computer mediation can decrease flaming because individuals focus their attention on the social context (and associated norms) rather than themselves."

[23] See P. W. Ewald, "Host-Parasite Relations, Vectors, and the Evolution of Disease Severity," *Annual Review of Ecology and Systematics*, 14, (November) 1983:pp. 465-485. [Available as: doi: 10.1146/annurev.es.14.110183.002341]; P. W. Ewald, "Evolution of Virulence," Infectious Disease Clinics of North America, 18, 2004: pp.1-15; P. W. Ewald, "Evolution of Infectious Disease, New York: Oxford University Press, 1994.

observation, it is found experimentally that among pathogens that when human subjects are exposed to a strain of the yellow fever virus that has cycled between humans and mosquitoes (its vector), the effects are much less severe than those produced by exposing a human subject to a strain of the virus carried by vectors that have not been in prior contact with humans. Like vector-borne pathogens, water-borne bacteria typically evolve more virulent strains than those that are transmitted by skin contacts with infected humans (and consequently loose their transport when their human hosts become immobilized. Thus, the virulence of diarrheal bacteria is observed to vary positively with extent to which they are waterborne: the proportion of severe strains among *Shigella*, for example, is found to be higher in geographic locales where there are opportunities for water-borne transmission.

The mechanism driving this differential behavior in microbial populations is the conditions of transport that permit, or prevent engagement of negative feedback from the pathogen's debilitating impacts upon their host. More virulent organisms' reproductive potentialities will be diminished relative to that of the others as a consequence of their destructive interactions with the hosts upon which they depend not only for nutrition, but for transport to other sites from which they and their progeny can draw sustaining resources – as does the *Mycobacterium tuberculosis* in consuming iron from the blood of its human host. Not so in the case of micro-pathogens (or "macro-parasites") that form no symbiotic "communities of dependence" with the medium that conveys them from one source of necessary resources to the next. The latter's replication therefore will not be impeded by genetic mutations and interactions with immobile hosts that result in the expression of greater virulence – which is to say, in correspondingly more debilitating or lethal effects upon the organisms with which they come into contact.

Now we consider the parable stretched in the other direction, by applying it to comprehend larger social systems – though not necessarily ones whose consequences for human societies are less damaging than the evolution of increasingly virulent microbial pathogens. This exercise can make its point by focusing on the outline resemblance between the foregoing stories of un-damped systems of social interaction in miniature, and the terms in which economists and economic policy analysts have come to explain the genesis and cataclysmic unfolding of the of the global banking and finance crisis of 2007-2008. Like the minor and transient episode of May 6th, the remarkable "bubble" that developed in U.S. house-prices, and the contemporaneous rise in degree of "leverage" attained by financial institution during this period, were phenomena "a long time in the making." So too as was the largely unnoticed evolution of a "shadow banking and financial system" that not only facilitated but drove these developments inexorably forward.

Crucial in that slower approach to the "edge of the manifold" beyond which lay monetary crisis and descent into widespread business failures and unemployment, was the gradually quickening emergence of novel financial intermediaries that operated outside the purview of supervisory agencies and regulatory codes that were supposed to reduce the prudential risks associated with moral hazard. The latter also were thought to be adequate restraints on the degree of systemic risks to which widespread illiquidity among banks and financial intermediaries could expose the "real" economy. These intermediaries deployed innovative financial instruments that soon connected new and long-established banks and investment houses alike in webs of contingent contractual claims, each one presumably of direct commercial benefit to the parties --if only because the novelty of their forms allow those using them to operate outside the domain of existing regulations and close governmental

oversight.[24] Credit default swaps (CDSs) ostensibly were a useful means of hedging against the risk of other companies defaulting on bonds that had been issued against collateralized debt obligations (CDOs), some of which were backed by mortgages of varying qualities – still another instrument that created an opaque structure of entanglements). But, CDSs turned out to have quite different system-level properties when they were created and deployed without restraint as a mode of betting on the likelihood of adverse business outcomes being suffered by third parties. They succeeded in enmeshing the counterparties in web of contracts characterized by increasingly correlated risks of mutual default, thereby obviously vitiating their ostensible usefulness as a means of diversifying against prudential and systemic risk.

The scope for banking and financial intermediaries to explore how they best could provide their customers and clients with these new and intriguingly lucrative "services" was expanding steadily in the U.S. throughout most of the 20th century's closing quarter – without drawing much general notice. This transformation of a previously staid and stolid branch of business, often disparaged by outsiders as "necessary but boring", had been accomplished incrementally by successive legislative steps that modified and blurred the sharp separation of commercial banking from investment banking activities – a separation that had been imposed in the U.S. by the passage of the Glass-Steagall Act during the Great Depression. The ultimate repeal of Glass-Steagall during the final year of the Clinton Administration therefore was largely symbolic, for by then most of the work of deconstruction had already been accomplished – certainly as far as concerned the operations of the large commercial banks.[25]

Yet, that symbolism heralded the onset of a still more permissive, deregulatory impulse in both the Congress and the agencies of the federal government under the administration of President Bush. The resolute "economic liberalism" animating Chairman Greenspan's resistance to exercising the supervisory powers and regulatory authority of the Federal Reserve System over innovative financial instruments and banking practices was paralleled by the recalcitrance of the Securities and Exchange Commission to subject investment banks to controls that would disadvantage "New York" in competition with the City of London and other international financial centers.

---

[24] This shared aspect of financial and technological innovations is developed further, below, in the concluding section.

[25] Charles C. Wang and Y. David Wang, in a recent paper ("Explaining the Glass-Steagall Inertia", Department of Economics, Stanford University July 20, 2009) show that that pressure for sweeping reform or repeal of the Glass Stegall Act in the early post-WWII era had been alleviated by the exploitation of loopholes in the 1933 statute. These limited the effectiveness of the Act's separation of investment banking from commercial banking, while leaving unaffected Senator Glass's long-sought objective of instituting federal bank deposit insurance. The effect of the former was to reduce the intensity of lobbying for the statute's repeal by the large commercial banks, whereas that of the latter maintained a congressional base of support for the Glass-Steagall Act's perpetuation – particularly among representatives of the interests of the country's less capitalized unit banks and the communities they served. As information technologies facilitating financial transactions and securities trading improved during the 60s to the 80s, however, changes in the competitive landscape marginalized U.S. commercial banks' relevance while weakening their ability to compete both at home and abroad. The wave of technological and regulatory changes that responded to that threat, being designed to improve the commercial banks' revenue growth and relative profitability, combined with repeated failures of more radical legislative proposals to find broad political support in the Congress during the 1980's and 1990s, allowed the Act's continuing survival in the statute-books. Wang and Wang argue that the eventual repeal of Glass-Steagall by the Gramm-Leach-Bliley Act of 1999 was the resultant of an essentially extraneous occurrence during the euphoria of the emerging Internet boom: the merger between Citicorp and Travelers Insurance in 1998, and the widespread perception that the latter would be a smashing business success, created sufficient political enthusiasm to overwhelm the factional conflicts that had doomed previous legislative efforts at reform.

Thus released from externally imposed checks on the impulse to seek greater profit through greater "leverage," positive feedback effects in the interactions among financial institutions were able to grow in power: they rushed to follow one another into new and riskier practices – finding reassurance about their comparative competitive vigor, and the illusion of "safety in numbers" in striving to maintain parity with their rivals and peers in their pursuit of ever more dubious lending practices. The magnitude of the destructive social consequences that can issue from allowing explosive economic processes of this sort to take hold of substantial claims on a "hyper-connected" system's assets are now only too familiar.

Closing reflections on the fraying nexus between financial instability and real economic growth in the 21st century

The excesses of the recent financial boom in the West and the ensuing global credit crises are not really new developments in the history of market economies, as Reinhart and Rogoff, and others before them have noticed.[26] One aspect, however, may be seen to be a signal of novel things to come while manifesting the familiar susceptibility of capitalist systems to this form of macro-instability. Nothing about that should be regarded as paradoxical: the discontinuous advent of novelties such as speciation in biological evolution is rooted in the continuity of mutant organisms' genetic endowments with that of their ancestors.

What we have recently experienced in the confined, high-frequency dynamical behavior of the U.S. national stock market on 6th May, and also in the property-finance-and-derivatives boom and crisis, are resultants of innovations that have affected and thereby disrupted both the sphere of technical affordances and the sphere of institutionalized regulatory structures. The synchronicity of the latter disruptions in each case, and in still others, is not entirely accidental.

Rapid advances in digital technologies, accelerated by the "connection-less" architecture of the Internet have enormously expanded possibilities of effecting conditions of hyper-connectivity in specific forms of transactions among human agents and machines ('things'). That modern society's augmented facilities in these respects have tended to outrun its capabilities for creating networked *communities* (in the particular sense of the term that has been invoked here) is neither a mere coincidence nor the product of concurrent independent processes one of which inherently moves at a slower pace than the other. Technological innovations engender positive private forces that contribute that the observed lag response in "governance," as has already been noticed. The business models of the operators of new electronic stock exchanges, like those of the small high-frequency trading firms that flocked into the national stock market, were aligned with and so worked to perpetuate the condition of high price volatility that normally is found in "thin" unregulated markets.

Thus, proximate reasons for the absence of mechanisms that would slow trading on the satellite electronic exchanges (paralleling the mandated cross-exchange "circuit breaker" regulations) can be found in the institutional consequences of the technical novelties that

[26] See C. M. Reinhart and K. S. Rogoff, *This Time Is Different: Eight Centuries of Financial Folly*. Princeton, NJ: Princeton University Press, 2010. This work is discussed in P. Krugman and R. Wells, "Our Giant Banking Crisis," The New York Review of Books, LVII(8), May 13, 2010:pp.11-13. Also, see, e.g., C.P. Kindleberger, *Manias, Panics and Crashes: A History of Financial Crises* 4th Edition. Hoboken, NJ: John Wiley & Sons., 2000. C.P. Kindleberger and R. Z. Aliber, *Manias, Panics and Crashes: A History of Financial Crises*, 6th Edition, forthcoming from the same publisher in 2010.

were driving alterations in the NMS's structure. Much of the volume of trading came to lie beyond the established ambit of the regulatory agencies, even though the registered equity and equity futures exchanges remained in other respects well within the jurisdictions of the SEC and the CFTC. Therefore, even if one leaves out of the picture the growing importance in the NMS of the unregistered "dark pools", there is a striking parallel between this aspect of the genesis of the May 6th "market break" and the failed regulatory supervision of the "shadow banking" system that had expanded quickly following the Glass-Steagall Act's repeal and the introduction of loosely regulated innovative financial instruments such as the CDOs and CDSs.

The riskiness of the CDOs would have been difficult for the rating agencies to establish accurately, even had they stronger incentives to do so, because the underlying packages of variegated mortgages were not like the usual assets against which corporate and government debentures conventionally were issued. As for the CDSs, being a form of insurance that deviated from conventional insurance contracts sufficiently for the firms writing them to make the case that they really were not insurance of the kind that called for the issuers to hold capital commensurate with the obligations they represented. Freed from the restraints that would otherwise have been placed upon their use within the regulatory jurisdiction of both the SEC and State commissions responsible for regulating the insurance industry, these devices for increasing "leverage" multiplied until they represented obligations far greater than even the most optimistic valuations of the assets they ostensibly were supposed to be insuring.

Situations of this sort occur repeatedly, in large part because established governance procedures and regulatory structures affecting transportation, communication or other network industries, and the contractual networks of banking and finance, historically have tended to be specified with reference to specific, widely-deployed technical or legal modes of transaction. They pertain not to generic functions, but specifically to the uses of railways, airplanes, and motor trucks, or telegraphy, telephones, radio and television systems, and to specific types of businesses (deposit banks, investment banks, and mutual saving associations) or to particular, classes of financial securities and contractual agreements. Legislative statutes addressing problems encountered with particular practices, especially technical practices, thus tend to confine themselves to dealing with the identified specific context of perceived dysfunctional outcomes. Rarely does social "rule-writing" for these domains tackle the generic goal, aiming to establish procedures that would achieve and maintain acceptable *standards of performance* for larger classes of socio-technical systems – including those who imminent emergence can be envisaged.

Rapid technological innovation, however, by creating novel and more effective means of connectivity, and permitting or inducing experimentation with new business models and other social practices (e.g., summoning "flash mobs") thus is able to have doubly disruptive effects, whether intended or not. In addition to their impacts upon the users of established technical practices, they induce the introduction of novel "forms" that more readily can escape the confines of the polity's legislative and administrative spaces. "Regulatory by-pass," permitting the formation of incompletely regulated or socially uncontrolled new businesses and technological systems can be an unplanned consequence of innovative activity. But the prospective profits to be reaped from successful regulatory by-pass also may be one of the strategic drivers of private investment in the deliberate design and implementation of system-disrupting innovations.

Modern societies, through their rush to create innovations in digital communication technologies and information processes, therefore may be creating a future multitude of emerging "networked exchanges" for which they continually will be failing to provide timely structures of governance, and passively encouraging fragmentation of previously integrated communities whose once functional interactions were sustainable but are becoming less so. This tendency might be tolerable as the regrettable cost of continuing economic dynamism. Indeed, some may be inclined to hail the consequences of the race in which innovation always stays a step ahead of governance and thus forms liberated spaces for the growth of profitable productive enterprises – portraying this process as the essential feature of "creative destruction" seen by Joseph Schumpeter to be inherent in capitalist economies' growth and development.

A judgement of that kind risks taking optimism to the limit of blindness, or denial of the possibility that we may no longer be living in the world that Schumpeter saw. Having highlighted salient historical continuities in the structural formations that are likely eventually to manifest themselves in transient market malfunctions and more profound financial crises that implicate real economies' abilities to fully utilize their productive capacities, it is necessary to conclude by drawing attention to a worrisome new departure in the workings of market capitalism that has become more obtrusive in recent years. Schumpeter's theory of economic development, one should recall, integrated real and monetary dynamics by assigning a central role to the financial accommodation of investments directed toward industries that would supply new products, or to the production of familiar goods and services by radically novel technologies. Credit-fuelled investment "manias" that involved experimentation with the use of novel financial instruments as well as reliance on familiar practices in banking and finance, were a recurring feature of the capital formation booms that figured in Schumpeter's schema; these movements were "excessive" in the sense that they were likely to outrun the expansion of consumption demand required to fully utilize the productive capacity that was being brought into existence.

The ensuing collapse of the final wave of real but nonetheless "speculative" investments, and the ending of the mania in a financial crash, thus are closely coupled events in the Schumpeterian business cycle model. Nevertheless, the manias' latent function -- a modern sociological construct that Schumpeter did not employ-- could be seen to have been that of effecting temporal coordination in the undertaking of many infrastructure investment projects whose "lumpiness" and complementary relationships made the commercial viability of each project interdependent with that of the others. None of those enterprises might have been successfully launched, but for the unreasoned optimism that had simultaneously animated their promoters and overwhelmed dispassionate recognition of the hazards of undertaking any one of them with without assurances of the others' being completed.

That the new capital structures erected in those episodes of over-investment were tangible and durable was a crucial analytical supposition of this analytical scheme, and it was an equally important premise for Schumpeter's sanguine normative stance regarding the cyclical instability of the macroeconomic growth process. He saw successive waves of temporarily "excessive" investment as having each left in its aftermath more efficient systems of production and distribution whose expanded capacity was available for more intensive exploitation during the ensuing phases of recovery. Indeed, their eventual utilization in the post-crisis phase of recovery was the source of the secular upward course of real income and wealth in those societies from the Industrial Revolution of the late 18th century onwards. In

this account of capitalist development, the positive trend is inextricably connected with the phenomenon of cyclical instability.

But it has become harder to see such a dynamic at work in the present century, where the investment manias associated with the "dot.com boom" and the "home mortgage and financial derivatives" bubble are popularly likened to the Tulip Mania of the Netherlands' Golden Age – an arcane topic in the annals of speculation schemes that until lately occupied the attentions of few people other than the social and economic historians studying pre-industrializing and financially interconnected regions of Northern Europe in the 17th century. The common thread tying the Dutch Tulip craze to its early 21st century counterparts -- and distinguishing all of them from the intervening historical succession of investment manias -- is that after they had "crashed" so little in the way of enduring productive assets was left behind.

More disturbing still, the latest episode would seem to have surpassed the turn-of-the-millennium boom in that dubious respect. The latter movement saw the excessive investment exuberance ignited by the Internet's advent become a scramble to subsidize the concentration of web traffic on competing websites and portals, and eventually animate investments in Internet's infrastructure layer. While it precipitated some spectacular mergers and acquisitions among telecommunications companies, the firms in that industry also were putting into place much new bandwidth predicated on expectations of continued hyper-exponential growth in web traffic. The mass of "dark fibre" that was left behind after 2001 certainly weighed heavily, for a time, upon the fortunes of the telecommunication corporations that owned it. But those real investments subsequently contributed substantially to the failure of the predicted congestion of the Internet to materialize during the years that followed. Viewed against that standard, the most recent episode of reckless mortgage investment in which global banking and financial institutions allowed were allowed, and allowed themselves to be caught up -- for the purposes of housing refinance and residential tract-building in arid regions remote from urban employment centers in California, Nevada and Arizona -- must be judged a largely sterile speculation.

If that is a harbinger of the future, it is a worrisome harbinger of a phase of sterile instability in the dynamics of capitalist "information societies" that focus their people's creative energies and imagination almost exclusively upon enabling new forms of "hyper-connected exchange networks," rather that on the challenge of creating "network communities" that can effectively govern the complex socio-technical systems they have built.

Figures

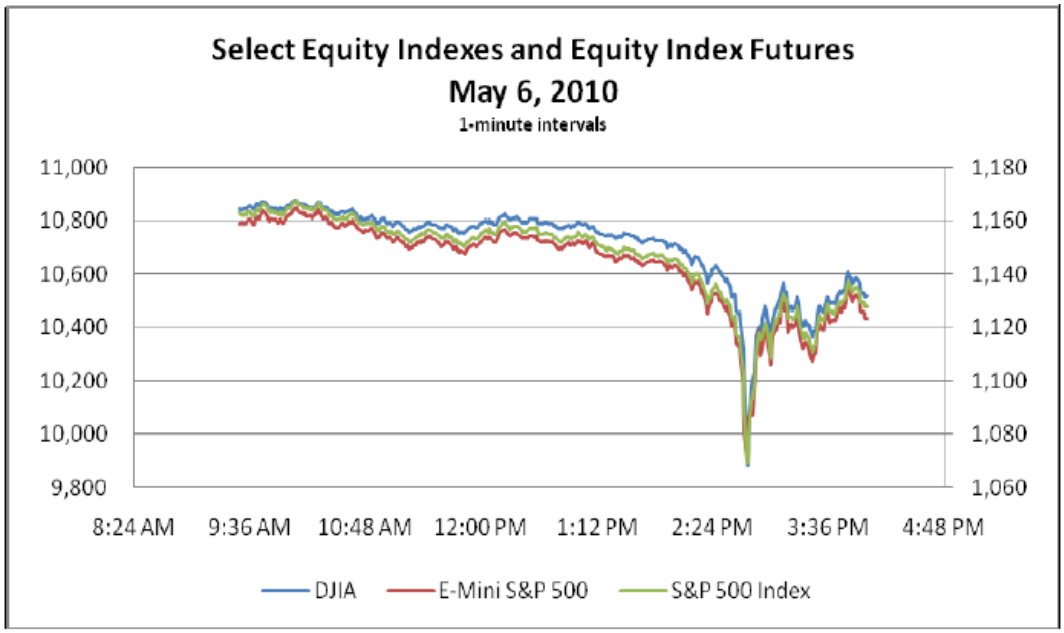


Figure 1

Source: http://www.sec.gov/sec-cftc-prelimreport.pdf:p. 17 (underlying data from *Bloomberg*)

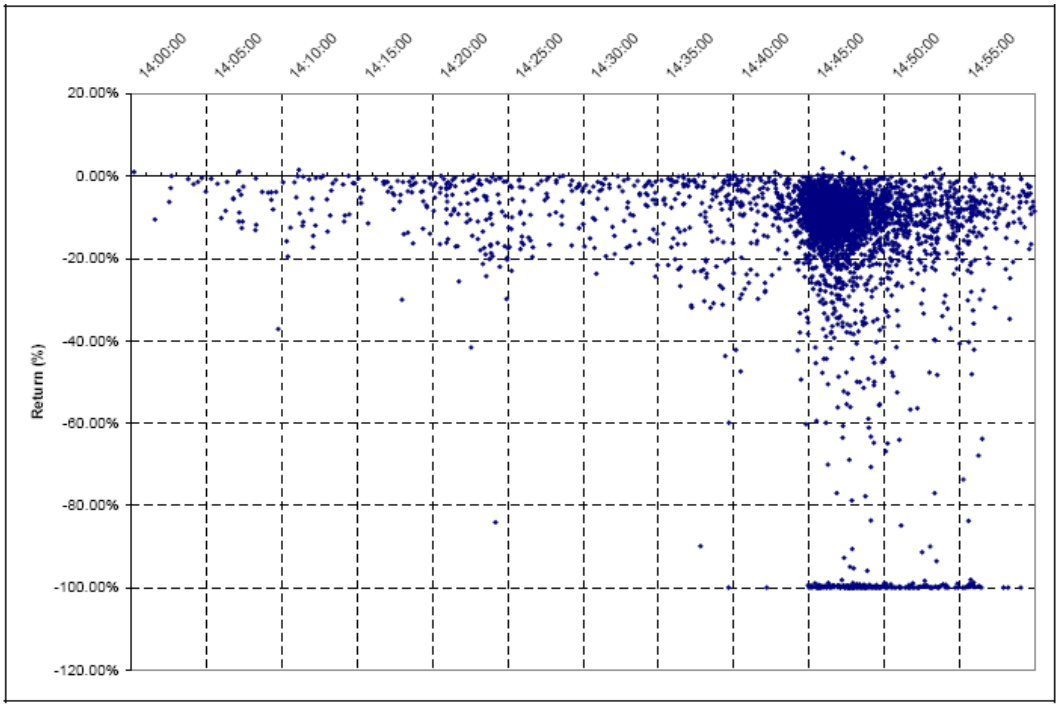


Figure 2

Source: http://www.sec.gov/sec-cftc-prelimreport.pdf:p. 24 (Underlying data from *Thompson Financial Datastream and NYSE Trades and Quotes*

Figure 2 depicts the timing of daily lows during the one-hour period from 2:00 p.m. to 3:00 p.m. on May 6. Each point represents the return from the May 5 close to the lowest transaction price on May 6, plotted against the time at which the transaction occurred. Daily lows not occurring during this one-hour interval are not depicted. The figure includes all equity securities (common and preferred) of corporate issuers, exchange-traded products, closed-end funds, and ADRs, traded on major U.S. exchanges, with a share price of more than $3.00 and a market capitalization of at least $10 million as of the May 5 close.

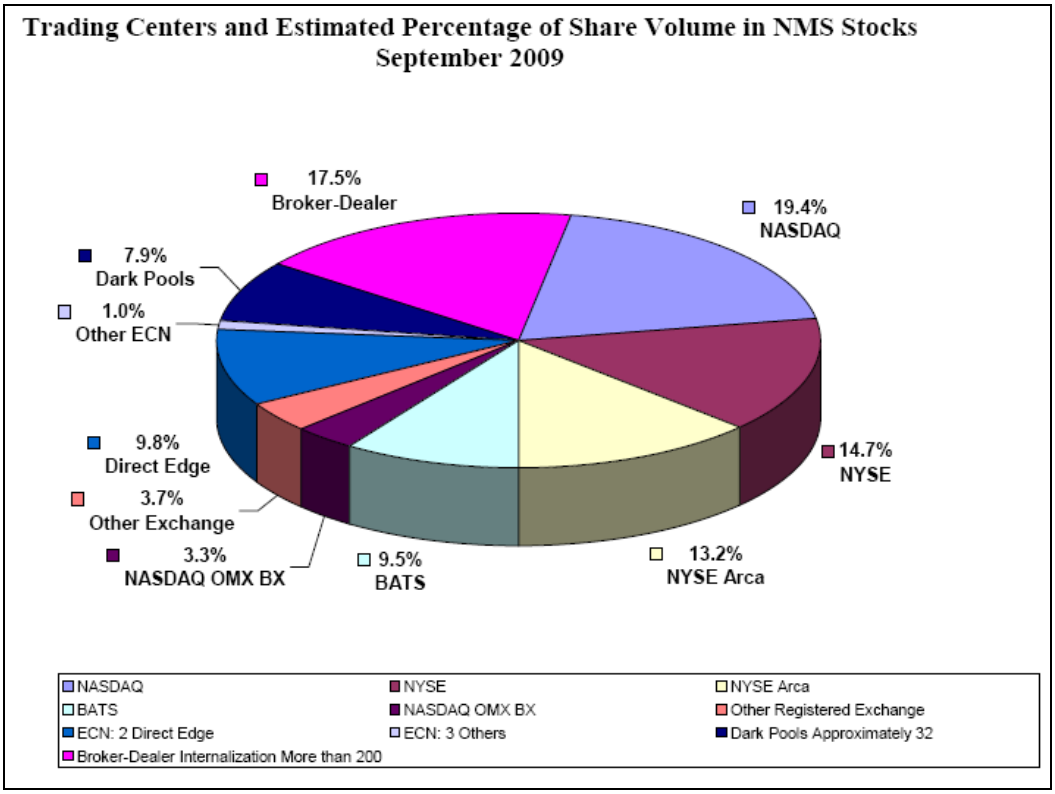


Figure 3

Source: http://www.sec.gov/sec-cftc-prelimreport.pdf Appendix A: p. A-3.

Acknowledgements

An early draft of this essay was presented at the *Imitatio* Economic Forum, in Paris at l'École Normale Supérieure on 13-14 June, 2010. I am grateful to *Imitatio* and its sponsor, the Thiel Foundation, for the opportunity afforded to expose the interpretations of events and their significance that I advanced on that occasion; and to the ensuing lively and penetrating discussion they received from participants in the Forum who approached these issues from the diverse perspectives of philosophy, heterodox economics, quantum physics and finance. The latter, it must be said, is far from being one of the fields of economics in which I am able to claim some degree of expertise. Even so, the subject of positive feedback and it implications for path dependence and other emergent properties of complex dynamical systems has long occupied my research and teaching, and lately my writings on economic policies addressing the challenges of global warming. The present (somewhat aberrant) research foray on my part was inspired by the conjunction of the events of last May 6th and my recent communications with Professor Jean-Pierre Dupuy, the director of research for *Imitatio*. Those conversations not only stirred recollections of the first coincidence of our interests -- two decades ago, at Stanford University -- in self-referentiality, reflexivity and self-reinforcing dynamics: it prompted my reading of Dupuy's paper "On the certainty of being surprised" (private correspondence, Paris 23 April, 2010). The latter in turn referred me to an article by Peter Thiel, "The Optimistic Thought Experiment," *Policy Review* (February-March 2008), which posed and sought to answer the question of how rational investors might best conduct their affairs when contemplating the likelihood of their economic system being overtaken by catastrophic events. Many of the that occasion observations that surface in these pages owe a good bit to their resonance with the issues treated in those two papers, although neither of the authors justly could be blamed for the approach I have taken, my reading of the events of May 6th, or the broader conclusions at which this essay arrives. The present version has benefited from Jean-Pierre Dupuy's perceptive suggestions concerning several passages in very preliminary draft, from the comments of Elizabeth Stone regarding the causes of the May 6th events, and discussions with W. Edward Steinmueller, Jean-Sebastien Bedo, and Daniel Hausmann about disruptive and potentially destructive effects of innovations enabling digital "hyper-connectivity". Lastly I must express special thanks to Charles Calomiris and Richard Sylla, for having responding quickly and in much detail to my requests for their "expert scouring" of the paper, which has spared readers of this version (and its author) the mistakes and omissions that marred the penultimate draft.

SUGGESTED CITATION:
Paul A. David, " May 6th -- Signals from a very brief but emblematic catastrophe on Wall Street", *real-world economics review*, issue no. 53, 26 June 2010, pp 2-27, http://www.paecon.net/PAEReview/issue53/David53.pdf

Beyond growth or beyond capitalism?

Richard Smith [USA]

***Abstract**: Recent publications have revived interest in Herman Daly's proposal for a Steady-State Economy. This paper argues, first, that the idea of a steady-state capitalism is based on untenable assumptions, starting with the assumption that growth is optional rather than built-into capitalism. I argue that irresistible and relentless pressures for growth are functions of the day-to-day requirements of capitalist reproduction in a competitive market, incumbent upon all but a few businesses, and that such pressures would prevail in any conceivable capitalism. Secondly, this paper takes issue with Professor Daly's thesis, which also underpins his SSE model, that capitalist efficiency and resource allocation is the best we can come up with. I argue that this belief is misplaced and incompatible with an ecological economy, and therefore it undermines Daly's own environmental goals. I conclude that since capitalist growth cannot be stopped, or even slowed, and since the market-driven growth is driving us toward collapse, ecological economists should abandon the fantasy of a steady-state capitalism and get on with the project figuring out what a post–capitalist economic democracy could look like.*

Under the headline "Economic Growth 'Cannot Continue'" the BBC on January 28, 2010 summarized a report issued by the New Economics Foundation (NEF) which asserts that "continuing economic growth is not possible if nations are to tackle climate change." The NEF says that "unprecedented and probably impossible" carbon reductions would be needed to hold temperature rises below 2°C (3.6°F) without which we face catastrophic global warming. "We urgently need to change our economy to live within its environmental budget," said NEF's policy director Andrew Simms, adding that "There is no global, environmental central bank to bail us out if we become ecologically bankrupt." [1] In *Growth Isn't Possible* Simms and his co-author Victoria Johnson reviewed all the existing proposed models for dealing with climate change and energy use including renewable, carbon capture and storage, nuclear, and even geo-engineering, and concluded that these are "potentially dangerous distractions from more human-scale solutions" and that there are "no magic bullets" to save us. The report concludes that even if we were to rapidly transition to an entirely clean energy -based economy, this would not suffice to save us because: "Globally, we are consuming nature's services – using resources and creating carbon emissions – 44 percent faster than nature can regenerate and reabsorb what we consume and the waste we produce. In other words . . . if the whole world wished to consume at the same rate it would require 3.4 planets like Earth." Given these facts and trends, Simms and Johnson argue, we have no choice but to bring average global growth to a halt (with sharp reductions in growth in the industrialized countries balanced by accelerated growth in the developing countries to approximate equity but tend toward stasis on balance) and to radically reconstruct the global economy to conform to "environmental thresholds, which include biodiversity and the finite availability of natural resources." The authors conclude that "a new macro-economic model is needed, one that allows the human population as a whole to thrive without having to rely on ultimately impossible, endless increases in consumption" and they point to Herman Daly's idea of a "Steady-State Economy" as their model. For a reaction to this report, the BBC asked Tom Clougherty, executive director of the Adam Smith Institute, a free-market think tank, for his response. Clougherty remarked that the NEF's report exhibited "a complete lack of understanding of economics . . . " [2]

[1] New Economic Foundation, *Growth Isn't Possible*, January 25, 2010 (London NEF, 2010) at http://www.neweconomics.org/publications/growth-isnt-possible.

[2] "Economic growth 'cannot continue' BBCnews Online, January 25, 2010 at http://news.bbc.co.uk/2/hi/science/nature/8478770.stm.

The NEF report comes on the heels of a new book published in December 2009 by Tim Jackson, Economics Commissioner on the Sustainable Development Commission, the UK government's independent advisor on sustainable development. In *Prosperity Without Growth* Jackson argues that our ever-increasing consumption adds little to human happiness, even impedes it, and is destroying our children's future. Jackson calls for a new vision of "prosperity without growth" and, like the NEF, points to Daly's Steady-State Economy as the best model.[3]

Now there is no doubt that the NEF is right that if CO^2 emissions continue to climb, catastrophic global warming will result. The NEF is also right that if there are no magic technofixes currently available, or in the foreseeable future, then the only way to stop global warming before it exceeds 2°C is to put the brakes on growth. But Tom Clougherty still has a point: pro-market but anti-growth economists don't understand capitalist economics. In rejecting the notion of a no-growth capitalism, Clougherty was just reaffirming the orthodox view of economists across the spectrum from Adam Smith to Karl Marx that growth is an iron law of capitalist development, that capitalism cannot exist without constant revolutionizing of productive forces, without constantly expanding markets, without ever-growing consumption of resources. [4] Indeed, it was precisely this market-propelled "motor" of economic development that for Karl Marx so sharply distinguished the capitalist mode of production from all previous historical modes of production like slavery or feudalism which contained no such in-built motor of development and so suffered repeatedly from stagnation, crises of *underproduction*, famine and collapse.[5] But of course *pace* the New Economics Foundation, the Adam Smith Institute believes that endless growth and ever-rising consumption are *good things.*

I. Why do capitalist economies grow?

Simms and Johnson begin by asking, "why do economies grow?" Their answer is that as a society we're "addicted" to growth.[6] Bill McKibben, in his Forward to Tim Jackson's book calls growth a "spell": "For a couple of hundred years, economic growth really was enchanting." But "the endless growth of material economies" threatens the underpinnings of our civilization. The "spell" can be broken and it is past time we did it.[7] Jackson says we can find a sustainable prosperity if we abandon the growth-obsessed, resource-intensive consumer economy, forget "keeping up with the Joneses," and "live more meaningful lives" by "downshifting" to consume less, find "meaningful work" and "revitalize the notion of public goods." "People can flourish without more stuff" he says.[8] For Jackson, Simms and Johnson

---

[3] Tim Jackson, *Prosperity Without Growth* (London: Earthscan, 2009).

[4] Smith's theorization of growth was rudimentary but clear. He believed that "division of labor is limited by the extent of the market." As division of labor increases output and sales (increases "the extent of the market"), this induces the possibility of further division and labor and thus further growth. Thus, Smith argued, growth was self-reinforcing as it exhibited increasing returns to scale. Adam Smith, *The Wealth of Nations* (various edns.) chaps. 1 and 3.

[5] For a more detailed discussion of Smith and Marx on these points, see my "The eco-suicidal economics of Adam Smith," *Capitalism Nature Socialism* 18.2 9 (June 2007) pp. 22-43.

[6] *Growth Isn't Possible*, pp. 8-15.

[7] *Prosperity Without Growth*, pp. xiii-xiv.

[8] Ibid., pp. 132, 150-151, 171, 193.

as for Daly, growth is seen to be entirely *subjective*, optional, not built into capitalist economies. So it can be dispensed with, exorcised, and capitalism can carry on in something like "stasis." So Jim Jackson tells us that in his vision of a "flourishing capitalism" the market would operate at a less frantic pace:

> Ecological investment calls up a different 'investment ecology.' Capital productivity will probably fall. Returns will be lower and delivered over longer timeframes. Though vital for ecological integrity, some investments may not generate returns in conventional monetary terms. Profitability – in the traditional sense – will be diminished. In a growth-based economy, this is deeply problematic. For an economy concerned with flourishing it needn't matter at all.[9]

Reading this, it's not hard to see why mainstream economists find the idea of a slow growth, let alone a no-growth capitalism, hard to take seriously. For a start, *under capitalism*, this would just be a recipe for mass unemployment among many other problems. A decade ago in the midst of the boom, Paul Krugman, writing in *The New York Times* wondered "if there isn't something a bit manic about the pace of getting and – especially—spending in *fin-de-siècle* America":

> But there is one very powerful argument that can be made on behalf of recent American consumerism: not that it is good for consumers, but that it has been good for producers. You see, spending may not produce happiness, but it does create jobs, and unemployment is very effective at creating misery. Better to have manic consumers American style, than depressive consumers of Japan . . . There is a strong element of rat race in America's consumer-led boom, but those rats racing in their cages are what keeps the wheels of commerce turning. And while it will be a shame if Americans continue to compete over who can own the most toys, the worst thing of all would be if the competition comes to a sudden halt.[10]

But then Paul Krugman is an economist. Ecological economists like to quote Kenneth Boulding who famously declared that *"Anyone who believes exponential growth can go on forever in a finite world is either a madman or an economist."* Boulding, Daly and their students say that economists like Krugman are living in denial if they think that growth can go on forever in a finite world. But Krugman and the mainstream could just as easily reply that Boulding and Daly are *themselves* living in denial if they think that capitalism can carry on without growing.

In what follows, I will argue that Herman Daly, Tim Jackson, Andrew Simms and the rest of the anti-growth school of ecological economists are right that we need a new macro-economic model that allows us to thrive without endless consumption. But they are wrong to think that this can be a capitalist economic model. I will try to show why ecologically suicidal growth is built into the nature of *any conceivable capitalism.* This means, I contend, that the project of a steady-state capitalism is impossible and a distraction from what I think ought to the highest priority for ecological economists today – which is to develop a broad conversation about what the lineaments of a post-capitalist ecological economy could look like. I'm going to start by stating three theses which I take to be fundamental principles and rules for reproduction that define any capitalism and shape the dynamics of capitalist economic development:

1. Producers are dependent upon the market: Capitalism is a mode of production in which specialized producers (corporations, companies, manufacturers, individual

[9] Ibid., p. 197.

[10] "Mon ey can't buy happiness. Er, can it?" *The New York Times*, June 1, 1999 Op-Ed page.

producers) produce some commodity for market but do not produce their own means of subsistence. Workers own no means of production, or insufficient means to enter into production on their own, and so have no choice but to sell their labor to the capitalists. Capitalists as a class possess a monopoly ownership of most of society's means of production but do not directly produce their own means of subsistence. So capitalists have to sell their commodities on the market to obtain money to obtain their own means of subsistence and to purchase new means of production and hire more labor, to re-enter production and carry on from year to year. So in a capitalist economy, everyone is dependent upon the market, compelled to sell in order to buy, to buy in order to sell to re-enter production and carry on.

2. Competition is the motor of economic development: When producers come to market they're not free to sell their particular commodity at whatever price they wish because they find other producers selling the same commodity. They therefore have to "meet or beat" the competition to sell their product and stay in business. Competition thus forces producers to *reinvest much of their profit* back into productivity-enhancing technologies and processes (instead of spending it on conspicuous consumption or warfare without developing the forces of production as ruling classes did for example under feudalism): Producers must constantly strive to *increase the efficiency* of their units of production by *cutting the cost of inputs,* seeking cheaper sources of raw materials and labor, by *bringing in more advanced labor-saving machinery and technology* to boost productivity, or by *increasing their scale of production* to take advantage of economies of scale, and in other ways, to *develop the forces of production.*

3. "Grow or die" is a law of survival in the marketplace: In the capitalist mode of production, most producers (there are some exceptions, which I will note below) have no choice but to live by the capitalist maxim "grow or die." First, as Adam Smith noted, the ever-increasing division of labor raises productivity and output, compelling producers to find more markets for this growing output. Secondly, competition compels producers to seek to expand their market share, to defend their position against competitors. Bigger is safer because, *ceteris paribus*, bigger producers can take advantage of economies of scale and can use their greater resources to invest in technological development, so can more effectively dominate markets. Marginal competitors tend to be crushed or bought out by larger firms (Chrysler, Volvo, etc.). Thirdly, the modern corporate form of ownership adds irresistible and unrelenting pressures to grow from owners (shareholders). Corporate CEOs do not have the freedom to choose not to grow or to subordinate profit-making to ecological concerns because they don't own their firms even if they own substantial shares. Corporations are owned by masses of shareholders. And the shareholders are not looking for "stasis"; they are looking to maximize portfolio gains, so they drive their CEOs forward.

In short, I maintain that the growth imperative is virtually a law of nature built-into in any conceivable capitalism. Corporations have no choice but to seek to grow. It is not "subjective." It is not just an "obsession" or a "spell." And it cannot be exorcised. Further, I maintain that these theses are uncontroversial, even completely obvious to mainstream economists across the ideological spectrum from Milton Friedman to Paul Krugman. But Herman Daly, Tim Jackson and the rest of the pro-market anti-growth school of ecological economists must *deny* these elementary capitalist rules for reproduction because their project for a "steady-state" eco-capitalism rests on the assumption that capitalist economic fundamentals are not immutable, that growth is "optional," and thus dispensable.

II. Ecological economics and the problem of growth

From the earliest efforts in the 1960s and 70s to bring ecological concerns to bear on capitalist economics and corporate practice beginning with the 1972 Club of Rome report *Limits to Growth* , mainstream pro-market eco-futurists, eco-capitalists, and ecological economists have tried to deal with the problem of capitalist growth in one of two ways: Either, with Herman Daly and his school, they imagined that capitalism could be reconstructed such that it would more or less stop growing *quantitatively* but continue to *develop* internally – much as, Daly suggested, we ourselves stop growing physically at adolescence but continue to develop our capabilities, intellect, skills, etc. Or, with Paul Hawken, Lester Brown and other "sustainable development" proponents, they imagined that capitalism could carry on growing more or less forever but that this growth could be rendered benign for the environment by forging an eco-entrepreneurial-led "green industrial revolution" and by introducing green subsidies and imposing carbon taxes, polluter pays penalties and the like to bring the rest of industry on board. Pro-growth or anti-growth, both approaches assume that capitalism is sufficiently malleable that capitalist fundamentals can be "inverted" such that corporations can, in one way or another, be induced to subordinate profit-making to "saving the earth." [11] But what unites both schools of thought is their *a priori* rejection of alternatives to capitalism, their rejection of any kind of economic planning or socialism. So Jonathan Porrit, former Chairman of the UK Sustainable Development Commission, ex-Green Party Co-chair and one-time Director of Friends of the Earth, spoke for the mainstream when he declared that "Logically, whether we like it or not, sustainability is therefore going to have to be delivered within an all-encompassing capitalist framework. We don't have time to wait for any big-picture ideological successor." [12] I will address the problems of the pro-growth "sustainable capitalist" models of Paul Hawken et al. in a separate paper. Here I am going to focus on the problems and contradictions of the pro-market anti-growth school whose foremost theorist is Professor Herman Daly.

III. Capitalism without growth?

In the 1970s and 80s, Herman Daly launched a broadside assault on the academic discipline of economics assailing its dogmatic and neo-totalitarian embrace of neoclassical economics and its willful blindness to our looming environmental crisis. In pathbreaking and widely influential books and articles Daly assailed the "stupor of economic discourse" by holding up to his colleagues what he called the "wild facts" of our ecological crisis: the growing hole in the ozone shield, the alarming evidence of rising CO^2 levels, the shocking rates of natural resource consumption, the frightening rates of extinction and loss of biodiversity and so on which mainstream economists ignored (and most continue to ignore to this day). The ecological crisis is caused, Daly argued by too much growth: "the scale of human activity relative to the biosphere has grown too large" and most especially, by ever-growing consumption in the advanced industrialized countries. Daly attacked the mainstream's "idolatrous" "religion of growth," its "growthmania," its "fetish" of limitless

[11] Eg. Hawken, *Ecological Commerce* (New York: HarperCollins, 1993) p. 11-13.

[12] *Capitalism as if the World Mattered* (London: Earthscan, 2005), p. 84.

consumption. [13] Daly's critique of the neoclassical defense of growth is probably the most devastating critique to come from within the profession.

But despite his "radical" break with the mainstream's fetish of growth, Daly did not at all break with his colleagues' fetish of the market organization of production, the capitalist market economy. On the contrary. His proposal for a Steady-State Economy was based, he said, "on impeccably respectable premises: private property, the free market, opposition to welfare bureaucracies and centralized control."[14] So in his Steady-State model, Daly embraces capitalism but he rejects the *consequences* of market-driven economic development, especially overconsumption and environmental destruction. Now one might reasonably ask, how can he have it both ways? Daly tries to get around this contradiction by *abstracting* from the day-to-day workings of capitalism, from the demands on corporate CEOs by shareholders, from the pressures of market competition, from the implications of a no-growth capitalism for employment, and so on, and talks instead about the economy at a highly abstract meta level. So Daly says that if we are not to overdrive our ecology, there must be a "macro, social decision" about limiting the scale of growth."[15] He says that "In my view," the industrialized countries, must "attain sustainability in the sense of a level of resource use that is both sufficient for a good life for its population and within the carrying capacity of the environment if generalized to the whole world. Population growth and production growth must not push us beyond the sustainable environmental capacities of resource regeneration and waste absorption. Therefore, once that point is reached, production and reproduction should be for replacement only. *Physical growth should cease, while qualitative improvement continues*." [16]

But how could there ever be a capitalist economy that does not grow *quantitatively*? For more than thirty years Daly has chanted his mantra of "development without growth" but he has yet to explain, in any concrete way, how an actual capitalist economy comprised of capitalists, investors, employees, and consumers could carry on from day to day in "stasis." Capitalist economies are, as noted above, comprised of individual producers, businesses and corporations, producing in competition with one another for sales on the market. Of course there are some, typically small, privately-owned businesses, or niche industries -- farms, restaurants, mom-and-pop stores, landlords, as well as larger sole ownerships, partnerships, and family-owned businesses which can, if they so choose, carry on producing and marketing more or less the same level of output year-in year-out so long as they don't face immediate competition -- because the owners of such businesses do not have to answer to other owners, to shareholders. Regulated public utilities comprise another category of enterprises that can also largely escape competitive pressures to grow because their sales, prices and profits are guaranteed and set in advance. But those are not most of the economy. Most of the economy is comprised of large corporations owned by investor-shareholders. Shareholders, even shareholders who are environmentally-minded professors investing via their TIAA-CREF accounts, are constantly seeking to maximize returns on investment. So they sensibly look to invest where they can make the highest return (these days, *any* return). This means that corporate CEOs do not have the freedom to choose to produce as much or little as they like, to make the same profits this year as last year. Instead, they face relentless

[13] *For the Common Good*, (Boston: Beacon, 1989), pp. 1-2; *Steady-State Economy* (Washington D.C.: Island Press, 1991), pp. 75, 100, 102, 103; *Beyond Growth* (Boston: Beacon 1996), pp. 10ff.

[14] *Steady-State Economy*, pp. 2, 54, 190-91.

[15] *Beyond Growth,* p. 16

[16] *Beyond Growth*, pp. 3,5 (my italics).

pressure to maximize profits, to make more profits this year than last year (or even last quarter), therefore to maximize sales, therefore *to grow quantitatively*. So automakers, for example, look to make a profit from every car they sell. They can do this either by increasing the rate of profit on each car they sell by i*ntensifying* production -- finding cheaper material inputs, cutting wages to lower labor costs or bringing in more efficient labor-saving technology. But they can't increase profits forever in this way. Competitors can find the same cheap inputs, the same new technology. And they can't lower wages below subsistence. So this avenue has limits. Or, they can try to maximize profits *extensively* -- by selling more cars. In practice of course carmakers do both but increasing sales is normally the main avenue of profit maximization because, as Adam Smith noted, returns are theoretically limited only by the extent of the market. So facing saturated markets at home, U.S. car makers look to Asia. The same goes for any other investor-owned corporation. They're all locked into the same competitive system. In the real world, therefore, few corporations can resist the relentless pressure to "grow sales," "grow the company," "expand market share"-- to *grow quantitatively*. The corporation that fails to outdo its past performance risks falling share value, stockholder flight, or worse. So Starbucks can't quench its investors thirst for profit with just coffee, even overpriced coffee, so its baristas push frappuccinos, mochaccinos, skinny cinnamon dolce lattes, CDs, movies – whatever it takes to keep profits rising. So Apple can't afford to take a breather after last year's huge success with its iPhone. Shareholders demand something new this year to propel stocks to new highs – et voilá: the "iPad" (whether you need it or not). Seen in this light, "growthmania" is hardly just a dogma, an ideology, a fetish. "Growthmania" is a rational and succinct expression of the day-to-day *requirements* of capitalist reproduction everywhere and in any conceivable capitalism.

And if economic pressures weren't sufficient to shape CEO behavior, CEOs are, in addition, legally obligated to maximize profits -- *and nothing else*. So when researching his book *The Corporation*, Canadian law professor Joel Bakan interviewed Milton Friedman on the subject of the "social responsibility" and the responsibilities of executives. Friedman, channeling Adam Smith, told him that corporations are good for society but corporations should not try to *do* good for society. Bakan summed up this discussion thusly: "Corporations are created by law and imbued with purpose by law. Law dictates what their directors and managers can do, what they cannot do, and what they must do. And, at least in the United States and other industrialized countries, the corporation, as created by law, most closely resembles Milton Friedman's ideal model of the institution: it compels executives to prioritize the interests of their companies and shareholders above all others and forbids them from being socially responsible – at least genuinely so." [17] In short, given unrelenting economic pressures and severe legal constraints, how could corporations adopt "stasis" as their maximand?

Why would anyone want a steady-state capitalism?

Of course there are times when capitalist economies do slow down, and grind along in a sort of stasis -- but that's even worse. Since the fall of 2008 when the world economy suddenly ground to a halt, we've been treated to a preview of what a no-growth stasis economy would look like *under capitalism*. It's not a pretty sight: capital destruction, mass unemployment, devastated communities, foreclosures, spreading poverty and homelessness, school closures, and environmental considerations shunted aside in the all-out effort to restore growth. That is "stasis" under capitalism. In one of his books, Daly wrote with some

[17] Joel Bakan, *The Corporation* (New York: Free Press, 2004) pp. 34-35.

exasperation, "must we [grow] beyond the optimum, just to keep up the momentum of growth for the sake of avoiding unemployment?" [18] Well, yes actually, because under capitalism workers don't have job security like tenured professors. This fact may partially explain why it is that, despite all the anti-growth books published since the 1970s, *there is no public support out there for a capitalist steady-state economy.* And why should there be? Why would anyone want a steady-state *capitalist* economy? Poll after poll shows that ordinary citizens want to see the environment cleaned up, want to see a stop to the pillage of the planet, the threat of destruction of their children's future. But *as workers* in a capitalist economy, "no growth" just means no jobs. If limits to growth are imposed, and some industries have to cut back, where would laid-off workers find re-employment? And if the economy does not continuously grow (*quantitatively)*, where would the jobs come from for the workers' children? Today, in the United States, there are said to be at least seven applicants for every available job. Where are those other six people going to find jobs if there is no growth? And this situation is far worse in the developing world where unemployment levels are off the charts. So throughout the world, *given capitalism*, the only hope for workers is more growth. As a recent headline in the satirical weekly *The Onion* ran: "Masses Clamor for New Bubble."

IV. Limiting 'scale'?

Daly says quite rightly that we need to reduce growth and consumption to save the humans. The way to do this he says is to limit the scale of "resource throughput." But what is "throughput?" Throughput, he tells us "is the flow beginning with raw materials inputs, followed by their conversion into commodities, and finally into waste outputs" [19] OK, but which resources and commodities? Do we need to limit production of meat, coal, oil, synthetic chemicals? How about Starbucks' frappuccinos, SUVs, Flat screen TVs? Ikea kitchens, jet flights to Europe, 12,000 square foot homes? Daly doesn't tell us. He doesn't think it's necessary to specify cuts in resource use or consumption because he believes the market is the best mechanism to make these micro decisions: *"Once the level of resource throughput is reduced to a sustainable level, the pattern of consumption will automatically adapt, thanks to the market. Let the market determine efficient allocation."* [20] Daly does see a role for government – to make the macro-decisions. He says that the government or "some democratically elected body" should set "controls" or "quotas" on consumption of particular resources. And the quotas, he says, *"must be low enough to prevent excessive pollution and ecological costs that fall on the present as on the future."* [21] But how could this work under capitalism?

For a start, those quotas would have to be awfully low for some industries like, say, commercial fishing, tropical logging, even lower for the most polluting industries like coal, and virtually zero for many chemicals – if we seriously want to protect present and future human generations not to mention other species. But how could any capitalist government deliberately reduce overall consumption to a "sustainable level" and/or impose steep cuts on particular industries? Reducing consumption means reducing production. But as we noted, under capitalism that just means recession, unemployment, falling revenues, or worse. So right now, no capitalist government on the planet is looking to do anything but *restore and*

[18] *Steady-State Economy*, p. 101.

[19] *Beyond Growth*, p. 28. Cf. *Steady-State Economy*, p. 36

[20] *Beyond Growth*, p. 17.

[21] *Steady-State Economy*, pp. 17, 53, 65.

accelerate growth. That's why the U.S. Congress killed the cap and trade bill, weak as it was. That's why at Copenhagen, no capitalist government was willing to sacrifice growth to save the environment. [22] But even during the recent long economic boom, no government would accept binding limits on emissions. So Copenhagen is only the latest in a long sorry string of failures: Bali, Nairobi, Rio, and all the way back to Kyoto in 1997. It would appear, therefore, that the chances of any capitalist government "reducing consumption to sustainable levels" are less than slim, at best.

Secondly, the ecological crisis we face is not only caused by the overall scale of production and consumption, it is *just as much* caused by the specific *irrational, inefficient, wasteful, and destructive nature of the capitalist market's "allocation of resources"* -- and equally, by the market's failure to allocate resources to things we *do* need. The problem is *what we produce, what we consume, what we dump, what we destroy*. So for example, NASA's Jim Hansen, the world's leading climate scientist, says that "Coal emissions must be phased out as rapidly as possible or global climate disasters will be a dead certainty." "My argument is that new coal-fired power plants must be stopped as a first step toward phasing out coal emissions [and phasing out our dependence on fossil fuels]." "Yes, most of the fossil fuels must be left in the ground. That is the explicit message that the science provides." [23] If we don't, we won't be able to contain global warming to within 2°Centigrade, and if we fail to do that, our goose is cooked

After global warming, global toxic chemical pollution is almost certainly the next greatest environmental threat we face. Scientists since Rachel Carson have warned that human survival and the survival of many other species is increasingly at risk because of the growing assault on our bodies and the environment from the tens of thousands of different kinds of toxic chemicals pumped, dumped, leached, sprayed, vented into the environment every year by the chemical industry, polluting factories and farms, power plants, and so forth.[24] In April 2010 the President's Cancer Panel issued a landmark 240 page report in which it said that "the true burden of environmentally induced cancers has been grossly underestimated" and strongly urged President Obama "to use the power of your office to remove the carcinogens and other toxins from our food, water, and air that needlessly increase health care costs, cripple our nation's productivity, and devastate American lives."[25] Except for lead, PCBs, DDT and a few others which have been banned or partially banned, toxic chemical pollution of all kinds has worsened dramatically in recent decades, all over the world, especially because of the flood of new synthetic chemicals in pesticides, plastics, fabrics, pharmaceuticals, cleaners, cosmetics, etc., and thus into our food, water and the air we breathe. The average American apple or strawberry is laced with pesticides, some of which did not exist in Rachael Carson's day.[26] America's favorite seafood, shrimp, "is a health

---

[22] See Jim Hansen's discussion of both Copenhagen and the U.S. climate bill in *Storms of My Grandchildren* (Bloomsbury, 12009), chapter 9.

[23] *Storms of my Grandchildren,* pp. 172, 178-9, and 236.

[24] Rachel Carson, *Silent Spring* (New York: Houghton Mifflin, 1962). Theo Colborn et al. *Our Stolen Future: Are We Threatening Our Fertility, Intelligence, and Survival?* (New York: Dutton, 1996).

[25] LaSalle D. Leffall, Jr. M.D. Chair et al. *Reducing Environmental Cancer Risk*, 2008-2009 Annual Report (U.S. Dept. of Health and Human Services, National Institutes of Health, National Cancer Institute, Washington D.C. April, 2010) at http://deainfo.nci.nih.gov/advisory/pcp/pcp08-09rpt/PCP_Report_08-09_508.pdf.

and environmental nightmare." [27] Chemicals used in rocket fuel and dry cleaning turn up regularly in baby formula.[28] In the United States, the increasing contamination of public water supplies all over the country has become a scandal and raised alarm.[29] Everywhere we turn, we're exposed to more and more toxins.[30] Today, some 80,000 chemicals are in use in the United States, barely two hundred of which have even been tested for toxicity to humans, and only a handful, actually banned. They're in our homes.[31] They're in our bodies.[32] And many are known to cause or are associated with birth defects, cancers, chronic illnesses and physical disorders, neurological disorders in children, hyperactivity and deficits in attention, developmental and reproductive problems in humans and animals – and these are on the rise around the world.

Given that we can't anticipate all the potential risks of new synthetic chemicals, and given the scale of the problem when hundreds of new chemicals are introduced every year and many released into the environment in huge quantities, even millions of pounds, scientists like Theo Colburn and her colleagues argue that "humans as a global community" need to reconsider the convenience of synthetic chemicals like endocrine-disrupting plastics, pesticides, and other products, "against the risk they entail" and consider a drastic reduction or even a phase-out:

> Phasing out hormone-disrupting chemicals should be just be the first step, in our view. We must then move to slow down the larger experiment with synthetic chemicals. This means first curtailing the introduction of thousands of new synthetic chemicals each year. It also means reducing the use of pesticides as much as possible . . ." *They confront us with the unavoidable question of whether to stop*

---

[26] Environmental Working Group, "A few bad apples: pesticides in your produce," April 2000 at http://www.ewg.org/reports/fewbadapples.

[27] Taras Grescoe, *Bottomfeeder: How to Eat Ethically in a World of Vanishing Seafood* (New York: Bloomsbury, 2008).

[28] Environmental Working Group (EWG) news release: "CDC: Rocket fuel chemical in most powdered infant formula," April 1, 2009 at http://www.ewg.org/node/27784.

[29] On the state of America's waters, see the *New York Times* series Toxics Waters by Charles Duhigg: "Clean water laws neglected, at a cost," September 13, 2009; "Debating just how much weed killer is safe in your water glass," August 23, 2009; "Health ills abound as farm runoff fouls wells," September 18, 2009; "Sewers at capacity, waste poisons waterways," November 23, 2009; "Millions in U.S. drink dirty water, records say," December 8, 2009; "That tap water is legal but may be unhealthy," December 17, 2009.

[30] Leslie Wayne, "Fight grows over labels on household cleaners," *New York Times, September 17, 2009.* Anjali Athavaley, "Kicking formaldehyde out of bed," *Wall Street Journal*, October 23, 2009. Joseph Pereira, "Protests spur stores to seek substitutes for vinyl in toys," *Wall Street Journal*, February 12, 2008.

[31] Leslie Kaufman and Gardiner Harris, "Environmental group reveals toxic chemicals in a range of consumer items," *New York Times,* September 17, 2009.

[32] Andrew C. Revkin, "Broad study finds lower level of old chemicals, but new trends are called worrying," *New York Times*, February 1, 2003. Anila Jacob, MD, et al. The Chemical Body Burden of Environmental Justice Leaders (Environmental Working Group, May 2009) at http://www.ewg.org/report/Pollution-in-5-Extraordinary-Women. Erika Schreder, *Earliest Exposures* (Washington Toxics Coalition, November 2009) at http://www.mnn.com/family/baby/blogs/study-finds-babies-are-exposed-to-toxic-chemicals-in-the-womb.
Bobbi Chase Wilding, Kathy Curtis, Kristen Welker-Hood, *Hazardous Chemicals in Health Care: a Snapshot of Chemicals in Doctors and Nurses* (Physicians for Social Responsibility, 2009) at http://www.psr.org/assets/pdfs/hazardous-chemicals-in-health-care.pdf.

> *manufacturing and releasing synthetic chemicals altogether*. There is not glib answer, no pat recommendation to offer. The time has come, however, to pause and finally ask the ethical questions that have been overlooked in the headlong rush of the twentieth century. Is it right to change Earth's atmosphere? Is it right to alter the chemical environment in the womb of every unborn child. It is imperative that humans as a global community give serious consideration to this question and begin a broad discussion that reaches far beyond the usual participants . . . [33]

So scientists are telling us that to save the humans we need to virtually shut down the coal industry, drastically reduce production of fossil fuels, and phase out many toxic chemicals as quickly as possible.[34] But, how can we do this under capitalism? Peabody Coal, Chevron Oil, Monsanto --these are huge companies which have sunk all their capital and trained thousands of skilled personnel to produce what they produce. How could they just write all that off and start over? How could they accept quotas that would force them to drastically reduce production, depress profits, or even close down -- and be responsible to their shareholders? As Milton Friedman said, "corporations are in business to make money, not save the world." Yet if corporations carry on with business as usual we're doomed. So what to do?

Lineaments of an ecological economy

If we're going to save the world, I would suggest that humanity is going to have to begin that "broad discussion" Theo Colborn proposed, with people across the whole of society and around the world to figure out how to redesign the economy. This could be the starting point of an eco-socialist economic democracy. For my part, I would suggest that an agenda for that discussion ought to include at least the following points: 1) We're going to have to find ways to put the brakes on out-of-control growth, even if it means drastically retrenching or shutting down coal companies, oil companies, chemical companies, auto companies, even whole economic sectors dedicated 100% to waste production like the disposable products industries. 2) We're going to have to radically restructure production to impose sharp limits on the production, to physically ration the use and consumption of all sorts of specific resources like coal, oil, gas, lumber, fish, oil, water, minerals, toxic chemicals, and many products made from them. Some products, like coal-fired power plants, toxic pesticides, diesel fuel, bottled water, junk food, should probably be phased out and banned altogether. 3) We're going to have to sharply increase investments in things society *does* need, like renewable energy, organic farming, public transit, public water systems, public health, quality schools for our children, and many other currently underfunded social and environmental needs. 4) We're going to have to do away with production that is geared to mindless consumerism and needless repetitive consumption and the industries that support them. Too many choices and too short a lifespan for products have socially and environmentally unbearable costs. We live on a small planet with limited resources. Others need those resources too, so we can't afford waste. 5) We're going to have to devise a rational approach to waste which means to minimize all waste, forbid the disposal of toxics of any sort, eliminate most if not all single-use products like disposable containers, wrappings, diapers, pens, cameras,etc., eliminate throwaway containers, enforce mandatory and systematic reuse of containers, packaging, recycling, composting, etc. 6) And, if we have to shut down polluting or wasteful industries

[33] *Our Stolen Future*, p. 246-47 (my italics).

[34] Keith Schneider, "Science academy recommends resumption of natural farming," *New York Times*, September 8, 1989.

then society is going to have to provide equivalent jobs, not just retraining or the unemployment line, for those all those displaced workers because, if we don't, there will be no social support for the drastic changes we need to make to ensure our survival.

Of course, the minute we start talking about shutting down the coal industry or pesticide producers, or forcing them to change, and "directing" resources into new industries, then we're talking about violating capitalists' "freedom" to produce and sell whatever they like, and consumer "free choice" to buy whatever we want and can afford. We would be *screwing up the market*. That's right. But that is exactly what we *have to do* because the rational efficient market is very efficiently liquidating every resource on the planet and wiping us out in the process. If we want to save ourselves and many other species, then we have to give up the freedom of capitalists to produce and sell as they please and consumers to buy whatever they like and can afford -- *in order to win the greater freedom* for humanity to breathe clean air, to have safe water to drink, to have safe food to eat, to live long and healthy lives free of toxics-induced diseases, to restore a forested, clean, safe, habitable planet we can pass on to our children. Such a democratic and ecological economy would of course be completely incompatible with capitalist property and capitalist organization of production. It would in fact require large-scale democratic planning of the entire economy.

V. Daly's misplaced faith in the market

Daly rejects any such interference with market organization of production because, like his mainstream colleagues, he believes that "the market is the most efficient institution we have come up with" and the only option we have.[35] He can say this because he subscribes to a capitalist conception of efficiency. Capitalist economists since Adam Smith have defined economic efficiency from the standpoint of the *production unit* – the factory, mill, mine, etc. (which, conveniently, the capitalists own): So in capitalist terms, the most efficient production method, technology, or economic system is the one that gets the most output from the least input, so produces the cheapest widgets and generates the most product/sales/wealth for a given investment of labor and raw materials. So Daly says the market "is wonderful for allocation." "Markets singlemindedly aim to serve allocative efficiency." [36] Since markets are such efficient allocators of resources, Daly believes that the role of the state should just be to "impose . . . quantitative limits on aggregate throughput . . . within which the market can safely function, *and then the market is left alone*." [37] But what exactly does this mean? Efficient for what end? Optimal for whom? And by leaving the corporations "alone" to maximize capitalist efficiency and optimality according to their interests, doesn't this just open the way to further social and environmental destruction, and thus to undermine Daly's social and environmental goals?

So if, for example, mountaintop removal is the most efficient method of getting the most coal out of the ground at the cheapest price (which it seems to be), but this system is based on horrific environmental destruction -- not unlike war – with exploding mountains flooding, burying and devastating whole communities, towns and counties, poisoning water supplies, wrecking local economies throughout Appalachia, and adding new health problems

[35] *Steady-State Economy*, p. 51. *For the Common Good*, pp. 14, 19,44-47; and *Beyond Growth*, pp. 13-14, 17.

[36] *Beyond Growth*, pp. 13, 32 (italics in original). Daly quoted in Porrit, op. cit., p. 78, (my italics); *For The Common Good,* p. 44-49.

[37] *Steady-State Economics*, pp. 88-89 (my italics).

to already burdened communities – while the very efficiency of production itself only serves to lower the cost of coal, promote increased coal combustion, and thus accelerate global warming -- what is so optimal and wonderful about this free market allocation of resources? Who cares if mountaintop removal is the most cost-efficient allocation of resources if what they're producing is killing us? [38]

If satellite-guided fishing trawlers, with nets the size of several football fields, are the most efficient means of maximizing the fish catch at the lowest possible price, but this strip-mining of the oceans has wiped out fishery after fishery, depleting many global fisheries to the point of extinction, even starving dolphins and seals, while wrecking the ocean bottoms, demolishing coral reefs and destroying deep water ecologies – what is optimal about this market allocation of resources from the standpoint of humanity, nature and future generations of fish – and fish eaters? [39]

If toxic chemical companies like Monsanto or Dupont manufacture Roundup or Benlate at the highest level of technical efficiency, in the cleanest factories, with the least waste going out the back door, what does this matter if the products they send out the front door and spray all over the planet are helping to extinguish life on earth? What kind of lunatic efficiency and optimality is this? [40]

If most of the American economy – from cars to appliances, from furniture to decoration, from fashion and cosmetics to throw away this and that – and all their supporting industries and services like advertising, credit cards, packaging and on and on -- are geared to *insatiable repetitive consumption*, to driving consumers to, as retailing analyst Victor Lebow described it back in the 1950s, "use up, wear out, burn up, and discard" perfectly good cars, TVs, clothes, phones, and buy something "new" and "up to date" even if what they have already is perfectly useful, even if the new replacement is trivially different, in an endless and ever-growing cycle of planned obsolescence and "forced consumption," what is optimal and efficient, let alone wonderful, about all this -- given the state of the world's depleted resources today? [41]

Now Herman Daly would never want to see the sorts of awful, irrational, wasteful and destructive free market resource allocations I've just described turn up in his Steady-State Economy. But aren't such corporate practices guaranteed to be there? Since in Daly's model of a steady-state capitalism, the government's role is only to set an upper limit on throughput consumption and then get out of the way, leaving the market "alone" and in charge, why would the market act any differently than it does right now?

Eco-socialist efficiency vs. capitalist efficiency

There is a place for efficiency in an ecological economy. After all, no one wants to waste labor or natural resources. But when, as under capitalism, the whole point of using resources efficiently is just to use the saved resources to produce even more commodities, to

[38] See eg. Tom Butler et al. eds., *Plundering Appalachia: The Tragedy of Mountaintop Removal Coal Mining* (San Rafael, CA: Palace Press Intl.: 2009) and, again, James Hansen op. cit.

[39] See eg. Michael Berrill, *The Plundered Seas* (San Francisco: Sierra Club, 1997)

[40] See Marie-Monique Robin, director, *The World According to Monsanto* (National Film Board of Canada et al., 2008) and her book of the same title by The New Press, 2009.

[41] The quoted phrases Victor Lebow were cited by Vance Packard in *The Waste Makers* (New York: David McKay, 1960) pp. 24,33.

accelerate the conversion of *even more* natural resources into products -- to be "used up, worn out, burned up, and discarded" so the cycle can begin all over again – capitalist efficiency turns into its opposite. In the 1860s, the English economist William Jevons famously observed that gains in technological efficiency – specifically, the more economical use of coal in engines doing mechanical work – actually increased the overall consumption of coal, iron, and other resources, rather than "saving" them, as many had hoped (because British officials were already growing concerned about running out of coal). As he wrote: "It is the very economy of its use which leads to its extensive consumption. . . [E]very . . . improvement of the engine, when effected, does but accelerate anew the consumption of coal." [42] This "rebound" or "backfire" was not a function of technological improvement *per se*. Under different social arrangements, if profit were not the goal of production, then such gains in efficiency could indeed save these natural resources for the benefit of society and future generations. But Jevons lived, and we live, under capitalism, and in this system, cheaper inputs only give producers greater incentive to "grow the market" by selling more product at lower prices to more consumers, and thus to push sales and profits still higher. So, ironically, the very capitalist efficiency and market organization of production that Daly celebrates just brings on the growth and further environmental destruction he so dreads.

But if we consider efficiency from the standpoint of *society and ecology*, including future as well as present generations, instead of just from the standpoint of the production unit, then the definition of efficiency is completely the opposite of market efficiency. So from a social-ecological perspective, it would be absurdly inefficient to waste resources producing goods and services we don't need, to produce goods designed to wear out or become obsolete as fast as possible -- just so we can do the same work all over again. Why would we want to do that? It would be so much more efficient and less wasteful to build cars, appliances, computers etc. to be as *durable* and *long lasting* as possible, to need as *few* "model" changes as necessary, to be as *upgradable* and *rebuildable* as possible – and take longer vacations. From society's standpoint, it would be not just inefficient, but suicidal to keep running coal-fired power plants that are killing us just because capital is sunk into them. It would be far less costly to society and the environment, for society to collectively absorb the cost of phasing these out and replacing these plants with renewable technologies we already have. From society's standpoint, it would be ruinous to contaminate the country's topsoil, pollute our public water supplies, and poison ourselves with an endless array of to toxic pesticides and other synthetic chemicals, just to produce corn or soybeans a few cents cheaper per bushel for a decade or so until the soil is completely exhausted and poisoned. If Monsanto can't afford to shut down its production of toxics, society could afford to close down those polluting plants and find other, better, employment for those talented and skilled but mis-allocated chemists and workers. And even if society decides that it needs some synthetic chemicals, to some extent, an eco- social chemical policy would start from the Precautionary Principle such as has already been elaborated by scientists, doctors, and grass-roots anti-toxics organizations like Safer Chemicals Healthy Families, which calls for safer substitutes and solutions, a phase-out of persistent bioaccumulative or highly toxic chemicals, publication of full right-to-know and rights of workers and communities to participate in decisions on

[42] William Stanley Jevons, *The Coal Question, 3rd edn.* (New York: Kelley, 1905) pp. 140-41, 152-53, cited in Blake Alcott, "Jevon's paradox, *Journal of Ecological Economics*, 54 (2005) p. 12. Even pro-industry Frances Cairncross notes that in the chemical industry "[t]hroughout the 1980s, companies like Dow and BASF steadily cut effluent per ton of product sold, but their final sales increased." So pollution increased even as they "cleaned up." *Costing the Earth* (London: The Economist Books Ltd., 1992) p. 269.

chemicals, publication of comprehensive safety data on all chemicals, and insistence on the immediate priority protection of communities and workers in the event of any threat.[43] And so on.

VI. Beyond capitalism

Daly and the anti-growth school are certainly right that we need to break out of the "iron cage of consumerism," "downshift" to a simpler life, find meaning and self-realization in promoting the common good instead of accumulating stuff. They call for an environmentally rational economy that conserves nature and resources for the benefit of our children and theirs, instead of consuming the whole planet right now. They call for a redistribution of wealth to those in need and for the construction of a society based not on possessive individualism but on a decent material sufficiency for everyone on the planet. And they call for a moral and spiritual transformation of our values away from materialism. Those are admirable goals. But we can't do any of those things under capitalism because under capitalism, we're all just rats racing in Paul Krugman's cages. We can't stop consuming more and more because if we stop racing, the system collapses into crisis. So it follows that we need a completely different kind of economic system, a non-capitalist economic system based on human needs, environmental needs, and a completely different value system, not on profit. Ecological economists from Herman Daly to Tim Jackson have called for a "new macro-economic model" a "new vision," a "new paradigm," a "new central organizing principle." But all they actually offer us are unworkable, warm and fuzzy capitalist utopias, with no plausible means of escaping the iron cage of consumerism or the "growthmania" of the market. Jonathon Porrit says that "like it or not" we have to try to find sustainability within a "capitalist framework" and forget about alternatives. But if the engine of capitalist growth and consumption can't be stopped, or even throttled back, and if the logic of capitalist efficiency and capitalist rationality is killing us, what choice to we have but to rethink the theory? Like it or not Jonathon, it's time to abandon the fantasy of a steady-state capitalism, go back to the drawing boards and come up with a *real* "new macro-economic model," a practical, workable post-capitalist ecological economy, an economy by the people, for the people, that is geared to production for need, not for profit. "Socialism?", "Economic democracy"? Call it what you like. But what other choice do we have? Either we save capitalism or we save ourselves. We can't save both.

SUGGESTED CITATION:
Richard Smith, "Beyond Growth or Beyond Capitalism?", *real-world economics review*, issue no. 53, 26 June 2010, pp. 28-42, http://www.paecon.net/PAEReview/issue53/Smith53.pdf

[43] See the Louisville Charter and its background papers at http://www.louisvillecharger.org/thecharter.shml; and the publications of Safer Chemicals Healthy Families at http://www.saferchemicals.org.

Happy talk and the stock market:

Why our situation is worse than many people think

David A. Westbrook*

Americans seem to view the stock market as the most important indicator of the nation's economic health. Equity trading activity is incessantly reported in almost real time; the Dow Jones Industrial Average (DJIA) shows up on the margins of newspapers and web pages that have little to do with business. In the present crisis, the "recovery" of the stock market is taken to indicate that the nation is on its way back to economic health, and by extension, that the structure of our economy is fundamentally sound, and regulatory reform presumably need only tinker around the edges.

Recent experience, however, does not encourage us to view stock price increases as good evidence of economic strength: the stock market was doing very well as we spiraled into the great recession. More generally, and for many reasons, stock indices are a poor reflection of our economic well-being: they tend toward "happy talk" that masks a more complicated and often disconcerting reality. And yet, the political will for financial market reform appears to fluctuate daily with the movement of the Dow and S&P 500.

In hope of cutting through some of the happy talk to a more honest policy discourse, it's worth exploring some of the ways that these indices are biased toward an unduly sunny view of our situation.

Time Slicing

By restarting the clock in January of 2009, after the near-panic selling of Fall 2008, we've established a misleading baseline for the DJIA and other indicators. The DJIA is still down almost a third from its peak in 2007, and roughly flat over the last decade (showing a loss if inflation is taken into account). The dramatic effects of time slicing may be easier to see in the context of a single company's stock. Imagine a company whose share price falls from a price of $100 to $1 in the final months of a given year. Now just imagine that the share price subsequently recovers to $10 (maybe the company successfully argues that it is too big to fail, and the government intervenes) in January of the following year. The lucky investor who (somehow foreseeing government intervention) invests at $1 will reap a return of 1000%. The stock is still down 90%. But we tend to focus on the 1000% return – the happy talk – not the 90% loss.

"Up" looks bigger than "Down"

Imagine a stock that falls from $100 a share to $50 a share, i.e., the stock has lost 50% of its value. Now imagine that the price rises to $75 a share. The stock has gained 50% of its value "back." Happy talk, even though the stock is actually still way down. Now imagine that the stock returns to 100, i.e., a 100% gain. More happy talk, but that's why 2009 was such a great year for the stock market, and why the administration and Fed clearly did such a "good job" in helping us to navigate the "tsunami."

Both time slicing and the rhetorical distortion of percentages can be illustrated with a significant, if unusually dramatic, example, AIG. On June 18th, 2007, AIG closed at

* Floyd H. & Hilda L. Hurst Faculty Scholar and Professor of Law, University at Buffalo, State University of New York. Westbrook's latest book is *Out of Crisis: Rethinking Our Capital Markets* (Paradigm 2009).

$1450/share. On March 9th, 2009, the share closed at $7, a loss of three orders of magnitude, but still less than 100%. With good timing, however, the investor who bought on March 9th would see the share price close on January 21, 2010, at $27.5, a roughly 400% profit in ten months, no doubt because AIG is a well run business.

Composition

The composition of an index changes over time. Companies that weaken sufficiently are no longer counted, but are replaced with other companies. The DJIA would look much different if it still included AIG, Citigroup, and General Motors. Thus, to compare the DJIA of, say, January 19, 2006 (12,566) with the DJIA of January 21, 2010 (10,389) is to compare two rather different sets of companies.

This is somewhat like fantasy football, if one could keep picking players as the season progresses: the DJIA reflects all of the stock price increases, but does not reflect the large decreases, of the companies that compose(d) it. Therefore the DJIA, like other indexes, has an upward skew, which conversely means that indicated losses can be significantly understated.

As the composition of the index changes, our ability to use the index to say how the market in question (and by extension, our society) is doing over time degrades, because it's not the same market. At best, an index indicates fluctuations of aggregate cost of capital to the kind of companies that happen to be on the index, i.e., an index is sort of a barometer of financial fashion. For example, during the internet mania, shares on the tech-heavy NASDAQ soared.

A market without friction

In order to match an index, an actual investor would have to buy and sell the right companies, at the right time, and incur transaction costs in doing so. None of this shows up in the index. More generally, indices do not attempt to represent the real world of investing; the index presents a "frictionless," and hence flattering, version of the market on which it is "reporting."

Demand noise

It is often assumed that demand for a stock indicates belief that the company is performing well. But when demand is high enough, and when alternative investments are not attractive (especially during crises, when asset classes fall in tandem), then "demand" may only indicate that the stock is "good enough for these hard times." Portfolio managers typically are under pressure to do something reasonably safe and profitable with the money under their care, and to do it now. This can be difficult. Leading into the present crisis, the world was awash in liquidity, and there literally were not enough investments to go 'round. At the worst of the crisis, investors bought U.S. government debt at negative yields. Such "investment" did not indicate belief that the US government was doing well; it indicated belief that there was no better place to park money. By the same token, and despite the crisis, today vast quantities of money need to be invested. The United States remains an intensely capitalized country: schools, hospitals, governments and retirement funds, and so forth – the structure of our society is mirrored in pools of capital, and that money has to go somewhere. Much of it goes into the stock market, especially now that interest rates are so low. Such demand for stock does not necessarily indicate that the companies in question are doing very well; it may well indicate reinvestment risk. (This can be observed with my own retirement

funds, such as they are.) Indeed, rather flat earnings, and more generally, sluggish growth (especially once inventory manipulations are factored out), and miserable employment numbers, suggest a very different interpretation of how businesses are doing. But we tell ourselves that the real economy lags the financial economy; hope is required.

Government bias

Much of the "recovery" of the stock market is due to the recovery of share prices in the financial sector. (At one point last summer, almost 40% of volume in the entire NYSE was due to trading in just four financial companies.) And why shouldn't these institutions make money? The government has ensured the operations of institutions deemed too big to fail. Much competition has been eliminated, as the companies picked to survive gobbled up those fated to die. The cost of capital has been lowered in various ways, including guaranteeing credit, purchasing assets outright, and allowing the use of trashy assets as collateral for low interest loans. With cheap money in hand, it is fairly easy to make money in financial markets. Institutional clients, who are in charge of vast amounts of money that must be managed, have certainly appreciated doing business in these difficult times with a company backed by something approaching the full faith and credit of the United States. Business has boomed; profits have risen.

Owning a financial institution that operates without much competition, a guaranteed spread, a gratefully captive customer base, and assurance of unlimited government funds if the going gets rough is a sweet proposition. Share prices have risen accordingly. While no doubt gratifying to those who own such companies, the rising prices of their stock hardly demonstrate that our financial policy has been a success. More generally, stock prices should and do reflect the fact that trillions of tax dollars have been spent upon some firms, including non-financial firms. However, such price increases do not necessarily indicate that our economy is healthy. In fact, the opposite seems more likely. Rather than the health of the economy, the index may well be reflecting the fact that it is a good thing to be paid, or guaranteed, by the government.

So equity indices are skewed upward. Nonetheless, and allowing for a degree of puffery, rising stock prices are a good thing, right? Not necessarily. It is true that rising stock prices can indicate a certain optimism, even momentum. As suggested by the example of massive government intervention with tax dollars, however, stock prices also may rise for reasons that are not good for anybody but shareholders. We should remember at least some of the basics of classical economics: in a competitive market, prices (and profits) tend to be driven down. This is good for consumers; this is the purpose of antitrust laws. In a competitive market, because profit margins are thin, stock prices should also be relatively low. Following this logic, widespread increases in stock price may suggest a correction of irrational downward trading, but may also indicate market power or government subsidy. The bottom line is that the investors' (including management's) interest in rising prices for their stock is completely understandable, but is not synonymous with a sound market, much less with the public interest in any broader sense.

The analytic distinction between the price of equity and the public good is crucial for thinking about government intervention during financial crises. If an investor's interest in the price of his stock is not the same thing as society's interest in a sound market, then we might draw a parallel distinction between the fortunes of the firms that currently dominate a business, and the business itself. For obvious example, we might differentiate between certain large *banks*, a small collection of very big companies, and *banking*, which is a valuable industry with many participants. Investors may be expected to care about the fate of

the particular banks in which they happen to be invested, but citizens, whether or not they are investors, should care about banking, and only incidentally about the fate of particular banks. By the same token, the government, presumably acting on behalf of all citizens, should care about banking, not about this or that bank. Although this distinction between banking and banks may seem clear enough in theory, in practice two administrations have conflated the need to intervene in the financial markets with the need to save particular, well-connected, institutions, and extended trillions of dollars in aid directly to such institutions.

The results of these highly targeted (and profitable) interventions have been disappointing. Credit remains tight, and the broader economy weak. The Federal deficit has soared to record heights, but over two years after the Fed started giving investment banks access to the discount window, efforts at reform have been modest. The financial industries have grown more concentrated, and remain vulnerable. Unemployment remains high, job creation low, foreclosures high . . . The government, in short, has used trillions of dollars to rescue some financial (and a few non-financial) institutions, but has not rescued banking, or financial markets more generally, much less the economy. We comfort ourselves by saying that no other form of intervention was possible, and the alternative was "another Depression." But the fact remains that we have very little – beyond stock price increases – to show for the expenditure of trillions. (Surely there were better ways to avoid another Depression.)

With little really good news, the stock market numbers have assumed even greater political significance. Improvements in the equity markets have been haled as "green shoots," signs of our economy's revival, that is, the government is on the right course, and as evidence that the system isn't that broken, and so little government action is required. While one must hope that a rising market expresses some renewed confidence in the real economy, a prudent skepticism is in order. As discussed above, the biases built into indices are overwhelmingly positive. This is not accidental. Our indices are essentially a way of defining and reporting on financial markets: Dow Jones, of Dow Jones Industrial Average fame, is a newspaper company. Treating the *Wall Street Journal* (another of Dow Jones' products) as revealed truth would hardly be prudent –financial journalism generally encourages increased participation in financial markets, much as the New York Review of Books is enthusiastic about reading. Unsurprisingly, financial journalism accentuates the positive for the financial industries and their investors – indices are merely a technical way of doing so. None of this should be confused with an impartial assessment of the economy's health.

While enthusiasm for asset prices is hardly new, the use of indexed data for "happy talk" has been particularly inane recently. The liberal media is, by and large, supportive of the administration, and so tries to emphasize the relatively strong performance of the stock market, while downplaying lackluster grown, bad unemployment, and downright shocking poverty numbers. The conservative media cheers for a stock market recovery and hopes to avoid increased regulation. Unfortunately, our financial markets, and our broader economy, have structural problems that those at either end of the political spectrum are unwilling to confront, presumably for fear of giving ground to their adversaries. Under such circumstances, interminable discussion of the stock market stands in for serious engagement with economic issues.

SUGGESTED CITATION:
David A. Westbrook, "Happy Talk and the Stock Market: Why Our Situation is Worse than Many People Think", *real-world economics review*, issue no. 53, 26 June 2010, pp. 43-46,
http://www.paecon.net/PAEReview/issue53/Westbrook53.pdf

The crisis in mainstream economics*

Geoffrey Harcourt [Cambridge University, UK]

In preparing this valedictory lecture I have been greatly helped by hearing and then reading Bob Rowthorn's speech to the King's Economists on 17 April; Paul Omerod's dissection of modern macroeconomics in the February 2010 issue of *21st Society*, the Journal of the Academy of the Social Sciences; Heinz Kurz's paper, "On the dismal state of a dismal what?", on the deficiencies of mainly Lucasian theory in the light of the current crisis events, together with his careful gathering together of Lucas's more outlandish and extraordinary claims for his approach and contributions and those of his tribe of admiring followers; especially when Keynes's contributions as they see them and which Keynes never claimed to have made, are used as his and their *numéraire*; Lance Taylor's "tome for our times", *Maynard's Revenge*, soon to be published by Harvard University Press; Robert Skidelsky's, *The Return of the Master* (2009); Joe Sliglitz's many criticisms of the extreme versions of modern theory which served to justify the Washington Consensus and its implications for universal policy proposals; Ben Friedman's review of recent books by John Cassidy and John Lancaster and Tony Judt's article, "Ill fares the land", in the latest issue of the *New York Review of Books*, April-May, 2010.

My title is pinched from John Hicks, *The Crisis in Keynesian Economics* (1974), his Yrjö Jahnsson Lectures; but let us also remember the apocryphal story of "Sunny Jim" Callaghan returning, sun-tanned, to strife-torn UK at the end of the 1970s and responding "Crisis, what crisis?".

In his book Hicks saw Keynesian economics in crisis on three fronts – the multiplier (because of the role of stocks); liquidity preference theory (because of complex portfolios of financial assets rather than a simple choice between money and bonds, coupled with concentration on the long-term rate of interest); and wages (Keynes's "failure" to provide an economic theory of the possible relationships between money wages, their levels and rates of change, and employment, its level and rate of change). So Hicks was tackling what he saw as deficiencies in a past theory when confronted with a (present day) reality. In my view Hicks rather overdid it because his view of Keynes was too much influenced by his own version of *The General Theory – IS/LM* – which by the 1970s dominated the profession rather than how Keynes himself had presented his theory in terms of the aggregate demand and supply functions. This provides yet another example of the tragedy that Lorie Tarshis's 1947 textbook did not dominate the post-war teaching of the economics of Keynes in the United Kingdom, the USA and elsewhere.[1]

There are similarities between this 1970s episode and what has happened in the last 30 years or more, now brought into sharp relief by the ongoing crisis in the capitalist world. Despite its great technical sophistication, in its conceptual essence, mainstream economics, now argued by its proponents to be increasingly converging on agreement and uniformity, is what Joan Robinson dubbed (as early as 1964) "pre-Keynesian theory after Keynes".

---

* A Valedictory Lecture given by G. C. Harcourt at SOAS on 12 May 2010. May I say how grateful I am to Jan Toporowski and SOAS for arranging the lecture (and the lunch) and to Sheila Dow and Jan Toporowski for their very kind introduction and closure respectively?

[1] For the story of why this was so, see Harcourt (1995).

Dominant figures in this transformation include Friedrich Hayek, Milton Friedman, Robert Lucas and Eugene Fama, with Lucas and Fama currently the patron saints of Chicago and modern macroeconomics (real and financial) and macroeconomists, including Michael Woodford and John Cochrane. Now that it is put to its first really challenging test, following the period of "the great moderation", let us examine whether its explanatory power and relevance have been found wanting.

Though there are several variants of modern macroeconomics, they all have their roots in (Irving) Fisherian Walrasian models of the process of accumulation in modern societies. In these, the consumer queen attempting to maximise her expected life time utility is the core actor and decision-maker, with all other actors and institutions subject to her whims and desires, especially within a competitive environment.

Fisher's basic vision and construct in theoretical terms was spelt out most fully and rigorously in the Arrow-Debreu model of general equilibrium. Subsequently, in the hands of Lucas and others, it was simplified in order to analyse the macroeconomic economy and to be the basis of stochastic general equilibrium models which at a practical level came more and more to serve policy makers in both central banks and treasuries. (At the same time Fisher's perceptive analysis of the consequences of debt deflation has largely been ignored.)

Concomitant with these developments was the development of the rational expectations hypothesis and its implications for views on how the economy functions. Though the rational expectations hypothesis by itself is no more than a hypothesis about expectations formation, something to be adopted until found wanting, when it is integrated with views of how the economy works, it becomes in its simplest and most stark form the proposition that the world may be viewed *as if* perfect competition and perfect foresight ruled in all markets, what Willhem Buiter aptly dubbed many years ago, "The macroeconomics of Dr Pangloss" (1980). For example, Lucas's policy ineffectiveness result follows not from rational expectations as such but from its use with a vertical aggregate supply curve. If a first-year student were to be asked what would be the impact on price and quantity of a rise in the demand curve in a market with a vertical supply curve, he/she would of course answer, price is higher, quantity is the same. As Joan Robinson once remarked (in another context), "After putting the rabbit into the hat in the full view of the audience, it does not seem necessary to make so much fuss about drawing it out again" (1966, 308).

Increasingly, in one strand of these developments, macroeconomic issues came to be analysed in terms of one representative agent models (Lorie Tarshis regarded this as the major heresy of (all) modern economics). This meant the rejection of any role for the fallacy of composition, a vital strand of the economics of Keynes. In turn this meant that the determination of the rate of interest could no longer be seen as the outcome of an uneasy truce at the margin between bulls and bears in financial markets; nor the rate of sustained inflation as establishing disappointed but not worsening aspirations between the capitalist accumulating and employing class and the wage-earning class. It also rejects another core Keynesian insight that the whole is often greater than the sum of the parts, now re-established in modern economics by Wynne Godley and Marc Lavoie's remarkable new book, *Monetary Economics* (2007).[2]

Another development, which on the face of it (and when examined more deeply, even by reading the original article), is the inappropriate use of Frank Ramsey's benevolent dictator

[2] May I pay here a heartfelt tribute to my old and much admired and loved friend, Wynne Godley, who died on 13th May 2010?

model to represent the essential workings of the economy. Ramsey's 1928 model of optimum saving was never so intended and it is salutary to reread or even read for the first time both it and its author's own evaluation of the article. When he submitted the article to the *Economic Journal*, he wrote to Keynes (28.6.1928). "Of course the whole thing is a waste of time … [It was distracting him from] a book on logic … [because] it [was] much easier to concentrate on than philosophy and the difficulties that arise rather [obsessed him]".

What of the new Keynesians? In his Marshall Lectures of some years ago (on a theorist looks at policy and a policy maker looks at theory), delivered when he was an advisor to President Clinton, Joe Stiglitz chose the new Keynesians as the modern development that most provided a rationale for Keynes-type results and policies. He also said that as a graduate student in Cambridge (UK) in the 1960s, he learnt most from the Cambridge Keynesians, especially Nicky Kaldor, and that it was their analysis and policies he drew on in his advice to Clinton. Nevertheless, he never once mentioned the Post-Keynesians, even though many of their ideas and insights were attributed by him to the other more fashionable "schools" he named.

The New Keynesians have made considerable contributions, not least when *internally* criticising mainstream macroeconomics – think Hahn and Solow (1995). Yet, even though their theories do produce Keynes-like results, these are traced back to imperfections in the working of market processes. This has the implication that the removal of such imperfections in real world happenings would usher in the world of Dr Pangloss – which is exactly the claims that the other strands make for their analyses. In particular, there is the major claim made that if competitive pressures were allowed freely to operate in all major markets – goods, labour, financial (national and international, long-term and short-term), foreign exchanges – for most of the time we would get near optimum workings of economies. Moreover, if there were exogenous shocks such institutional set ups would be the best way of allowing systems to adjust and quickly remove their effects. The high priest of these views is/was Alan Greenspan (though his mentor is the appalling Ayn Rand).

As is now well known, in the laundry basket at Tilton, Keynes's drafts of the differences between the cooperative, neutral and entrepreneur economies were discovered after volumes XIII and XIV of the *Collected Writings* had been published, resulting in a supplementary volume, XXIX. These contrasts figured prominently in Keynes's lectures at Cambridge prior to the publication of *The General Theory* and their omission was the event that most surprised and distressed Tarshis (who had been to the lectures) when he read *The General Theory*. Why? Because he thought them the best base on which to place Keynes's critique of the (non-) operation of Say's Law in an unregulated capitalist economy, see Harcourt (1995, 1246). Rather like an Evangelical Christian asking "Brother, are you saved?", Joan Robinson would ask what could or could not be determined directly by the actors most critically affected by the decision – the money wage or the real wage (the money wage, of course)? And Lorie Tarshis' litmus paper test for acceptance intellectually was which way does causation run – from the goods market to the labour market or the other way around? The entrepreneur economy was one of Keynes' ways of showing how and why monetary and financial matters must be integrated with real factors from the start of the analysis of a monetary production economy.

It is this insight that is missing from virtually all strands of modern mainstream theory. They have not taken on board the fundamental criticism that Frank Hahn made many years ago (and Colin Rogers (2006) has revived recently in his criticism of Cochrane and Woodford), that there is no place for anything recognisable as money in general equilibrium

models. Thus both the cycle and the trend (which, in post-Keynesian analysis, are regarded as indissoluble), in mainstream theory are taken to be operating independently of monetary and financial matters. Real business cycle theory, which has some similarities with 1930s Hayekian conjectures (though Hayek certainly understood about the role of money) is exactly what it calls itself. And Lucas argues that the world operates for most of the time near enough to a steady-state rate of growth as to use the latter as a short cut in his explanations. "The balanced growth path will be a good approximation to any actual path "most of the time" … exactly the reason why the balanced path is interesting to us", Lucas (1988, 11). Jean-Baptiste Kaldor of 1950s-1960s vintage could not have put it better.

Then there is the hoary old question of the essential difference between risk and uncertainty, so essential to the economics of Keynes, how they can be analysed and what their effects are on systemic behaviour. Bob Rowthorn (2010) makes the point that while microeconomic theories of risk have been systematically and usefully advanced, systemic spill over effects have been badly neglected in the analysis of a modern world characterised by wide-ranging networks of financial markets. Thus the possible malfunctionings of the latter and the feedbacks into real sectors, so prominent in the happenings of the last three years and more, have not been analysed in any fundamental ways in mainstream economics. Omerod (2010) adds to this insight by pointing out that the analysis of risk in financial markets rests on the assumption that underlying distributions usually approximate to the normal distribution and in particular "fat tails" are assumed away. He directs us to a long-established literature on the possible chaotic effects of the existence of fat tail distributions and relates this to recent happenings.

I have mentioned Lucas's arguments concerning the applicability of steady-state analysis usually associated with the theory of the long period (though Post-Keynesians would say, *correctly*, that it should not ever be thought to apply to the actual long run). But there is an element of wanting to have it both ways present when at the same time the short period and the long period are collapsed into one and all markets are taken to be clearing all the time. The Lucasians also argue that if we do not start from an assumption of utility maximising individuals, we are inevitably guilty of *ad hockery*. In doing so they ignore Kaldor's critique, that this approach leads to begging the question about the world being observed, that the observed happenings must have been thrown up by such behaviour in an equilibrating setting in which the factors responsible for existence (unique or otherwise) may be separated from those responsible for stability (local and global). Though path-dependence and multiple equilibria have taken increasingly conspicuous roles in the most sophisticated versions of mainstream theory, they have had little systematic effect on its more down-to-earth and more used versions. Certainly the possibility of cumulative causation processes operating in many important markets and indeed in systems as a whole is rarely if ever discussed, let alone countenanced.

Nor is there ever much use made of the distinction between the investment goods sector and the consumption goods sector, or between the capitalist class (all three) and the wage-earning class when analysing processes at work in modern capitalism. A person from Mars would be hard put to find much if any resemblance between the theory with which he/she was presented and the world with which he/she/it was confronted.

To sum up, there is a crisis in mainstream economics, in the teaching of it and in its application to theory and policy. For, by and large, it neither makes sense of what has happened or of what should and could be done about it. I would not go anywhere as near as far as Joan Robinson in "Spring cleaning" (1980, 1985) – scrap the lot and start again. Rather

I am somewhere in between discerning a crisis and "Sunny Jim's" supposed response. We do need a thorough rethink and regrouping in order to back up the tentative measures being taken at the moment to tackle the present crisis (they are very much a curate's egg approach, often more bad than good in parts), to better explain how our modern world functions and malfunctions and what may be done about it by people of goodwill who are humane, progressive and pragmatically practical. Immodestly, I hope I may be regarded a member of this band.

References

Buiter, W. H. (1980), "The macroeconomics of Dr Pangloss: A critical survey of the new classical macroeconomics", *Economic Journal* 90, 34-50.

Cassidy, John (2009), *How Markets Fail: The Logic of Economic Calamities*, Farrar, Straus and Giraux.

Friedman, Benjamin M. (2010) "Two roads to financial catastrophe", *New York Review of Books*, LVII, April-May, 27-29.

Godley, Wynne and Marc Lavoie (2007), *Monetary Economics*. *An Integrated Approach to Credit, Money, Income, Production and Wealth*, Houndmills, Basingstoke, Hants: Palgrave Macmillan.

Hahn, F. H. and R. M. Solow (1995), *A Critical Essay on Modern Macroeconomic Theory*, Oxford: Basil Blackwell.

Harcourt, G. C. (1995), "Lorie Tarshis 1911-1993: An appreciation", *Economic Journal*, 105, 1244-55.

Hicks, John (1974), *The Crisis in Keynesian Economics*, Oxford: Basil Blackwell.

Judt, Tony (2010), "Ill fares the land", *New York Review of Books*, April-May, 17-19.

Kurz, Heinz (2010), "On the dismal state of a dismal what?", mimeo, University of Graz.

Lancaster, John (2009), *I.O.U.: Why everyone Owes Everyone and No One Can Pay*, Simon and Schuster.

Lucas, R. E. (1988), "On the mechanics of economic development", *Journal of Monetary Economics*, 22, 3-42.

Omerod, Paul (2010), "The current crisis and the culpability of macroeconomic theory", *21st Society*. *Journal of the Academy of Social Sciences*, 5, 5-18.

Ramsey, F. P. (1928), "A mathematical theory of saving", *Economic Journal*, 38, 543-59.

Robinson, Joan (1964), "Pre-Keynesian theory after Keynes", *Australian Economic Papers*, 3, 25-35.

Robinson, Joan (1966), "Comment on Samuelson and Modigliani", *Review of Economic Studies*, 33, 307-8.

Robinson, Joan (1980, 1985), "Spring cleaning", mimeo, Cambridge; published as "The theory of normal prices and the reconstruction of economic theory" in G. R. Fiewel (ed.) (1985), *Issues in Contemporary Macroeconomics and Distribution*, London: Macmillan.

Rogers, Colin (2006), "Doing without money: A critical assessment of Woodford's analysis", *Cambridge Journal of Economics*, 30, 293-306.

Rowthorn, R. E. (2010), "King's Speech 17 April 2010", mimeo, Cambridge.

Skidelsky, Robert (2009), *Keynes. The Return of the Master*, London: Allen Lane. An imprint of Penguin Books.

Tarshis, Lorie (1947), *The Elements of Economics. An Introduction to the Theory of Price and Employment*, Boston: Houghton Mifflin, 2nd Edition, 1966.

Taylor, Lance (2010), *Maynard's Revenge*, Cambridge, Mass: Harvard University Press.

SUGGESTED CITATION:
Geoffrey Harcourt, "The crisis in mainstream economics", *real-world economics review*, issue no. 53, 26 June 2010, pp. 47-51, http://www.paecon.net/PAEReview/issue53/Harcourt53.pdf

Copeland on money as electricity

Anne Mayhew [Emerita, University of Tennessee, USA]

Bear with me while I summon the spirit of a too-often forgotten economist, one Morris Copeland, who did more than any other to create the flow-of-funds accounts that inform so much of modern economic macro analysis of the real world economy (Copeland 1952).[1] This summons from the dusty peace of library stacks seems essential In the face of the currently wildly misleading and ill-informed public debate over the perils of deficit spending by governments, a debate that threatens the world with more years of economic hardship and worse. Before turning to Copeland, consider again the need for better information and understanding.

Readers of this journal will certainly agree that public debate over the role of government finance in combating recessionary forces in our global economic has been seriously impoverished by the way in which textbooks and some financial journalists simplify or, to put it more accurately, talk down and obfuscate when describing processes of intersectoral finance in modern economies. When the public is led to believe that governments finance deficits by printing money, it is little wonder that there is anger and worry over the probable outcome.

People who are already angry over job losses, public corruption, stagnant wages and the many other problems that have been well documented over recent years in the U.S. and elsewhere are justified in wondering where those printed dollars go. The helicopters of the monetarist metaphor for money creation have not been seen dropping printed dollars over their neighborhoods. It is little wonder that there have been calls for greater transparency on the part of the FED and belt tightening by the government that is alleged to determine the drop zones of those helicopters.

Of course, readers of this journal know that money-dropping helicopters do not exist. Indeed, most economists know that the textbook and journalistic version of FED actions is not what happens in the real world economy, but the need for simplification at the introductory level and in the world of television sound bites and quick-read newspapers has served to perpetuate a quasi-official textbook view of money that derives from a long history in which money was a real stuff such as gold, or pieces of official paper, for which various forms of "credit" were adequate substitutes in good financial times, but from which sensible people ran in times of panic and distress. In this treatment it is important to maintain—as most textbooks do—a hard and fast distinction between credit, which is created in the ongoing interactions of a variety of public and private agents in the economy, and money, which exists as a result of the self-regulating market mechanism itself (as in an idealized gold standard) or through policy determination (as through the actions of the Federal Reserve System).

This distinction between money and credit makes little or no sense in our world of electronic payments, but, even more importantly, it also serves to obscure public discourse. At some point in their training, most economists learn, as do major participants in the financial sector and the journalists who write about them, that governments do not actually "print" money and most learn to rely for actual understanding of economic conditions on the "flow-of-funds" accounts (that Copeland pioneered and that are now compiled and published quarterly

[1] This brief paper is based on a longer paper on Copeland and his model presented at an interdisciplinary conference on "Money and Metaphors" held at the University of Virginia in October 2009. I will be glad to so send an electronic version of the draft of that paper. For more on Copeland and his contribution see Millar (1990) and Rutherford (2002).

by the FRS) rather than upon measure of the quasi-official money supply of textbook fame. In this passage to a more sophisticated, or perhaps simply truer, understanding of how the economy works, there is recognition that the idea of a central bank determined "money supply" may be as fictional as the Wizard of Oz or Santa Claus. Indeed Frederic S. Mishkin, author of a widely used textbook in which the quasi-official version of money is the basic model, has also written (with co-author Arturo Estrella) that at least since 1979, there is no empirical evidence to show that either M1 or M2 (the standard measures of the quasi-officially defined money supply) serve as "information variables" (Mishkin 2007; Estrella and Mishkin 2007). In other words, the standard measures of a "money supply" do not tell us what monetary policy is, which certainly implies that monetary authorities may not have the control that the textbooks allege. If we cannot conclude that the standard measure of money supply reveals the intention of the FED, then either the FED is lying (as many in the U.S. are inclined to believe), or something is wrong with the way we are thinking about the powers of the FED. However, my intention here is not to revisit and bring forward into the 21st century Keynes' liquidity trap, or to argue about whether, a la Friedman and Schwartz (1963), the fault lies with a timid and ineffectual FED. Rather my goal is to say that the quasi-official measure of money does not tell us much that is useful and that the textbook way of thinking about the relationship of money, finance, and spending is deeply flawed.

Alternative models are part of a well-established Post Keynesian and heterodox tradition in economics. However, this literature has not made its way into the standard texts and even into serious journalism, as evidenced by continued references to the helicopters of monetarist fame. One reason for this, I suggest, is that the alternatives proposed have lacked a simplifying metaphor suitable for easy presentation. And, this is where Morris Copeland comes in. Back in 1952, when Copeland, a Cornell University economist who had worked closely with Wesley Mitchell at the early NBER, published the first estimates of money flows in the United States, he not only provided the basis for the now widely used flow-of-fund accounts . He also proposed replacing the commonly used hydraulic model of the money supply—one in which money is thought of as a stock of water that flows through pipes or is held in reservoirs by banks—with a model of money as electricity. [2] In Copeland's electrical analogy, the reservoir become batteries and the conduits are wires. In a schematic picture, credit poles for each sector of the economy are fed by earnings from the "real" sector, which in the case of households would mean mostly wages and salaries, *and* by financial transactions, including transfers and, most importantly, by borrowing from other sectors. Borrowing sectors acquire funds and lending sectors acquire assets in the form of promises to pay.

Copeland's model of money flows integrated the "real" and the "monetary" sectors of the economy and was based on the accounts generated by businesses, households, and governments in the normal course of carrying on transactions in the real world economy. There are other crucial advantages as well. First, because there are no significant lags in the flows within and between sectors, there is no need to speak of a separate supply of and demand for money. In the case of electricity we speak of the ability of a system to generate

[2] The NBER began the great task of construction "national" or "social" accounts with the National Income and Product Accounts, which continue in the classical tradition of a sharp separation between the "real economy" of goods and services, in which money serves only as a measuring device. This left the monetary sector to be modeled using the more traditional tools of supply and demand. Wesley Mitchell and his disciple, Copeland, knew, however, that money had in many ways become as "real" as "real" goods for households and certainly for business firms. With the ever wider use or mortgages, credit purchases, and retirement accounts this has become ever more the case in recent decades. Copeland sought to revise national accounting to take this reality into account, but was not successful in persuading economists to introduce flow-of-funds accounts along with NIPA accounting into our textbooks. For more on this see Mayhew (2009).

electricity and of circuit breakers, but not of a supply that lurks behind the walls and is somehow there independently of the demand for electricity. Or, to put it differently and in the more precise language used by accountants, the crediting of an account automatically generates a debit in like amount to another account and the two are simultaneous, just as is the switching on of an electrical appliance and the draw upon electrical capacity. This is the process of modern economies and the electrical analogy suggested by Copeland closely approximates it.

A further advantage of Copeland's model, and one that makes it a nice supplement to the important work of Godley and Lavoie (2007), is that it stresses the discretionary power of the different actors in the modern economy. Using the more familiar hydraulic model, in which money is thought of as a stock of water, a stock of liquidity; we are required to think that something outside the financial grid causes an increase or decrease in the quantity. Money, as traditionally thought of, can flow rapidly or slowly or simply sit in a reservoir. This causes no end of confusion about velocity but it also requires that we think that central banks have a power that is belied by empirical evidence. Rather than argue about velocity, it is so much easier to say that in the dreadful period of 1929 to 1933, households and business firms didn't, for a variety of reasons, turn on the switches that cause money, thought of as electricity, to flow. Partly they did not do so because banks did not give them access to switches to people unlikely to be able to repay, but, even where they households and businesses did have access , they didn't use it and money, rather than sitting idle, just did not get generated.

Bring that idea forward, and note that if money is thought of as water and as a finite substance, then to spend it you have to take it away from someone else or, as newspapers and textbooks have it, increase something that is being printed and handed out, perhaps by helicopters or maybe by politicians. Of course, this is not what happens. As governments spend, they rely on transfers from households and businesses when taxes and other levies are paid, but they also issue promises to pay (either to themselves as Randy Wray [2006] would emphasize) or to others. Copeland's model has the enormous advantage of bringing this aspect of deficit finance to the fore. For every increase in government debt there is an increase in the holdings of assets by some sector of the economy.

The call for greater transparency on the part of the FED and of governments in general has been a major part of the political response to the current and ongoing recession. But, such transparency is going to be hard to achieve if you have to rely on hypothetical helicopters and money drops or measures of money supply that tell you little about what the FED's intentions may be. Actually, the FED itself, using a modified form of Copeland's money flow accounts, does a pretty good job of explaining what it does and what happens in the economy in its flow-of-funds accounting data. But, there is little to nothing in the textbooks that most of our students read that equips them to read these accounts and there is certainly no model or simple metaphor that helps in understanding the multiple sources of discretionary decision-making that determine macroeconomic outcomes. Much less are there explanations that are easily accessible to a television audience.

Copeland's electric metaphor could be the tool that is required to give us the transparency that could greatly improve public discourse. To adopt Copeland's model as both accounting tool and as guiding metaphor will require that we give up the idea of an all powerful if frequently bumbling FED and accept that ours is a complex and global electrical grid of money access, with very real circuit breakers that can in fact be managed in the public interest. To think in this way will require that we abandon inherited notions about money, abandon a lot of old debates that have engaged economists, and abandon the simple satisfaction of blaming central bankers. What we would gain in understanding and

transparency would be enormous and, besides, the ghost of Morris Copeland would, I think, be mightily pleased.

References

Copeland, M. *A Study of Moneyflows in the United States,* Publications of the National Bureau of Economic Research No. 54. New York: National Bureau of Economic Research, 1952.

Estrella, A. and Mishkin, F.S. "Is There a Role of Monetary Aggregates in the Conduct of Monetary Policy?" in *Monetary Policy Strategy,* edited by F.S. Mishkin. Cambridge, Mass: The MIT Press, 2007.

Friedman, M. and Anna Schwartz, *A Monetary History of the United States 1867-1960.* Princeton: Princeton University Press, 1963.

Godley, W. and Marc Lavoie, *Monetary Economics: An Integrated Approach to Credit, Money, Income, Production and Wealth.* New York: Palgrave Macmillan, 2007.

Mayhew, A. "Money as Electricity," a paper prepared for "After the Crash, Beyond Liquidity," A Conference on Money and Metaphors held at the University of Virginia, October, 2009.

Millar, J.R. "In Memoriam: Morris A. Copeland, 1895-1989," *American Journal of Economics and Sociology* 49, no. 1 (1990): 45-46..

Mishkin, F.S. *The Economics of Money, Banking, and Financial Markets* (8th edition). Addison-Wesley, 2007.

Rutherford, M. "Morris A. Copeland: A Case Study in the History of Institutional Economics," *Journal of the History of Economic Thought* 24 (2002): 261-90.

Wray. L. R. "Banking, Finance, and Money: A Socioeconomic Approach," Working Paper No. 459. Annandale-on-Hudson, NY: The Levy Economics Institute, 2006.

SUGGESTED CITATION:
Anne Mayhew, "Copeland on money as electricity", *real-world economics review*, issue no. 53, 26 June 2010, pp. 52-55, http://www.paecon.net/PAEReview/issue53/.Mayhew53.pdf

Debunking the theory of the firm—a chronology

Steve Keen and Russell Standish [Universities of Western Sydney and New South Wales, Australia]

A personal introduction by Steve Keen

I have been an economics renegade for almost 40 years, and for most of that time I have had to tolerate neoclassical economists either dismissing my work or ignoring it. Since the Global Financial Collapse began, that has ceased to be the case. Neoclassical economists, although hardly likely to become born-again Post Keynesians, are chastened and to some degree humbled by their inability to anticipate this crisis. Moreover, it's hard for them to hide behind the "Black Swans" defence when the speaker is someone credited with having predicted the crisis before it happened.

So that's all and good; one might even think that the seeds have been laid to finally achieve a true reform of economics, to make it empirical rather than a priori.

However, one also notices a worrying trend among neoclassicals today. True, after having had their macroeconomics lambasted and having been shown how Minsky's analysis makes sense of what they can't understand, they are often willing to admit that neoclassical macroeconomics is a shambles. No-one, they will say, defends rational expectations anymore. But then they assert: *at least neoclassical microeconomics is sound.*

This is patently absurd – especially since neoclassical macro was built largely by subverting its "Keynesian" predecessor on the grounds that it "did not have good microfoundations", and then casting macroeconomics as applied neoclassical microeconomics. However, despite this, neoclassicals still cling to the purity of their vision of the utility maximising consumer on one side of the market, and the profit-maximising firm on the other.

But absurd though it may be, it is an echo of what happened when Keynes tried to bring economics into the real world eighty years ago. Led by Hicks, Samuelson and the like, the neoclassicals dragged the profession back into fantasy via their vision of a beautiful microeconomics.

For society's sake we need to prevent this intellectually reactionary behavior from being repeated after this crisis. A dispassionate analysis of neoclassical microeconomics shows that it is absurd on its own grounds – that, to coin a phrase "neoclassical microeconomics lacks good microeconomic foundations". But the superficial elegance of the theory remains seductive, and when nations have got beyond this crisis, the same seductive superficiality could give rise to a neoclassical resurgence.

It therefore seems extremely important to emphasize and demonstrate again that their microeconomics is irreparably flawed. This was the task of my book *Debunking Economics*: to point out that, for example "downward-sloping market demand curves" don't slope downwards unless there is just one consumer and one commodity (the Sonnenschein-Mantel-Debreu conditions, treated properly as a proof by contradiction that individual demand curves can't be aggregated), that a marginal product of capital can't be defined (Sraffa's critique), and so on.

In most of *Debunking Economics*, I simply repackaged established critiques by previous authors – I stood on the shoulders of the giants that neoclassical economics ignored. But there was one part of the theory that, when I began the book, simply appeared irrelevant rather than analytically false: the theory of the firm. From my own experience as a young

believer in mainstream economics, I realised how powerful that vision of intersecting supply and demand curves in a competitive, efficient market is, but all I could do, I thought, was parody its irrelevance rather than analytically dismiss it.

Then I spotted a flaw that, to my knowledge at the time, had not been noted before: that the assumptions of a downward sloping market demand curve and a horizontal firm demand curve in the perfectly competitive model were mutually incompatible.

I since have discovered that I wasn't the first to point this out – Stigler, of all people, had done it in 1957. But I added a proof that what neoclassical economics calls profit-maximising behaviour – equating marginal cost and marginal revenue – provably does not maximise profits.

Further analysis found many other flaws in the superficially seductive "logic" of the neoclassical theory of the firm. In many ways, the flaws in this crucial part of neoclassical microeconomics are worse, and more easily proven, than those in consumer theory or capital theory or the like.

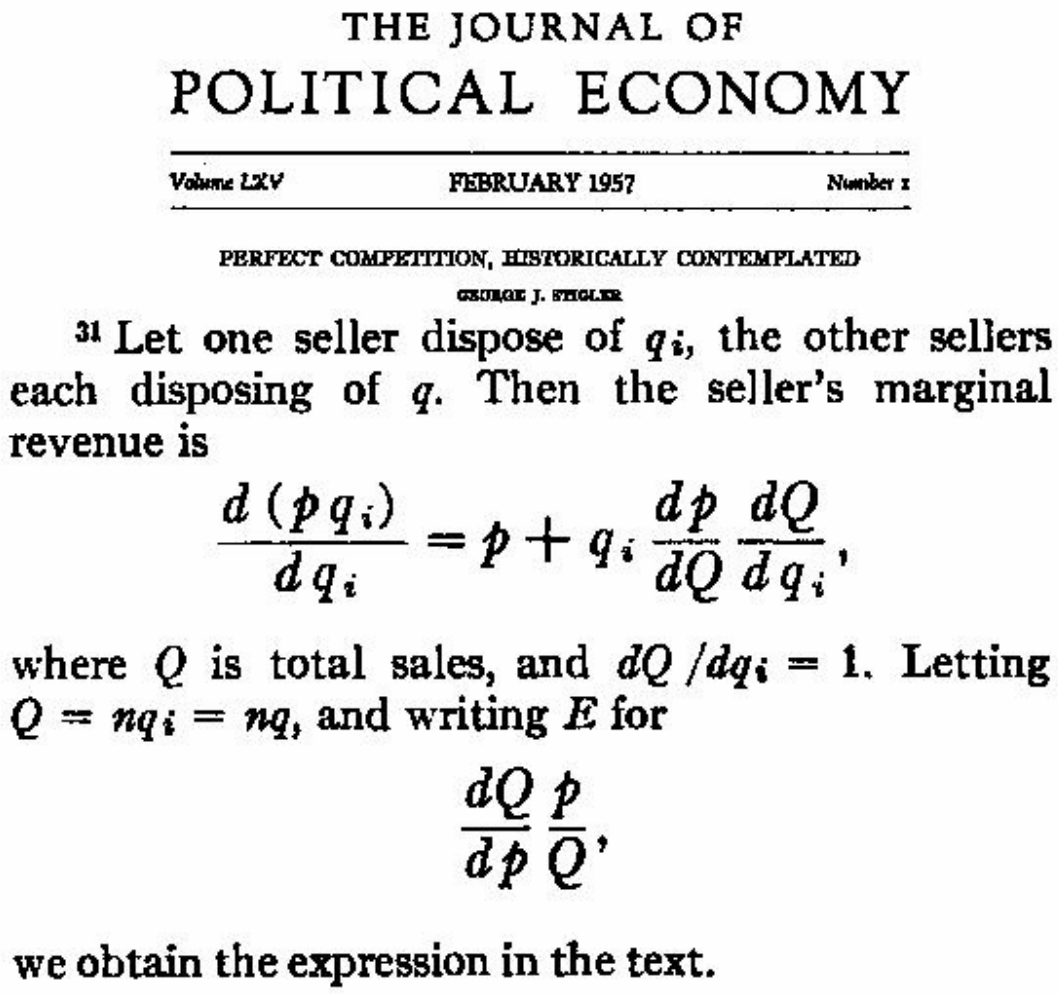
Of course, attempts to get this analysis published in mainstream economic journals prior to 2008 predictably failed; my various papers appeared in non-mainstream books and journals, and even the journal of interdisciplinary physics, *Physica A*.

I was content to leave them there and focus on my main interest today, of extending Minsky's work to understand the financial crisis. But then Australia's Treasury drafted a new tax whose foundation was the intellectually flawed vision of a competitive market, with firms facing a horizontal demand curve, and profit-maximising by equating marginal cost and marginal revenue.

So I decided to put together this compendium of all the reasons why this widely believed theory is nonsense. I start with the simplest disproofs that I have developed, and progress right through to a critique of Cournot-Nash theory. It's heavy going, but I fear that unless we really drive a stake through the heart of this vampire logic called neoclassical economics, it will rise again and lead us back to ignorance once we forget the Global Financial Collapse, as our predecessors forgot the Great Depression.

Stigler 1957

The proposition that Keen had thought original in *Debunking Economics*—that, under conditions of “atomism”, the slope of the demand curve facing the individual competitive firm was the same as the slope of the market demand curve—had in fact been made in 1957 by that arch defender of neoclassicism, George Stigler, and in a leading neoclassical journal: *The Journal of Political Economy* (Stigler 1957—see Figure 1).



THE JOURNAL OF
POLITICAL ECONOMY

Volume LXV FEBRUARY 1957 Number 1

PERFECT COMPETITION, HISTORICALLY CONTEMPLATED

GEORGE J. STIGLER

[31] Let one seller dispose of q_i, the other sellers each disposing of q. Then the seller's marginal revenue is

$$\frac{d(pq_i)}{dq_i} = p + q_i \frac{dp}{dQ}\frac{dQ}{dq_i},$$

where Q is total sales, and $dQ/dq_i = 1$. Letting $Q = nq_i = nq$, and writing E for

$$\frac{dQ}{dp}\frac{p}{Q},$$

we obtain the expression in the text.

Figure 1: Stigler 1957

Stigler's simple application of the chain rule showed that the underlying assumption of the Marshallian model—atomism, that firms in a competitive industry do not react strategically to the hypothetical actions of other firms—is incompatible with each firm facing a horizontal demand curve. In an n firm industry where the output of the i^{th} firm is q_i, this assumption, means that $\frac{\partial q_i}{\partial q_j} = 0 \ \forall i \neq j$. As a result, $\frac{dQ}{dq_i} = 1$, and hence $\frac{dP}{dq_i} = \frac{dP}{dQ}$:

$$\begin{aligned}\frac{dP}{dq_i} &= \frac{dP}{dQ}\frac{dQ}{dq_i} \\ &= \frac{dP}{dQ}\left(\sum_{j=1}^{n}\frac{\partial q_j}{\partial q_i}\right) \\ &= \frac{dP}{dQ}\left(\frac{\partial q_i}{\partial q_i}+\sum_{j\neq i}^{n}\frac{\partial q_j}{\partial q_i}\right) \\ &= \frac{dP}{dQ}\left(1+\sum_{j\neq i}^{n}0\right) \\ &= \frac{dP}{dQ}\end{aligned} \qquad (0.1)$$

It is thus impossible for the market demand function $P(Q)$ (where $Q=\sum_{i=1}^{n} q_i$) to have the dual properties that $P'(Q)<0$ and $P'(q_i)=0$—and Stigler had shown this in 1957! Yet the claim that the market demand curve is negatively sloped, while the individual perfectly competitive firm faces a horizontal demand curve, has graced the opening chapters of every economics textbook published in the last half century.

Mendacity in education—another personal observation

One of my motivations for writing *Debunking Economics* was my belief that an education in economics was mendacious. I had in mind the failure to note the Cambridge Controversy arguments when discussing the concept of an aggregate production function (see Chapter 6 of Debunking Economics), or the avoidance of the Sonnenschein-Mantel-Debreu conditions when deriving a market demand curve from the aggregation of individual ones (Chapter 2).

When I discussed these issues with any of the minority of neoclassical economists who were themselves aware of those critiques, the even smaller minority who did not dismiss them outright would raise the pedagogic defense of difficulty. These topics are complex, and require an advanced knowledge, not only of economics, but of mathematics. Better to give new students a simple introduction—well behaved aggregate production functions, nice downward sloping market demand curves, and so on—and cover the nuances when they have more knowledge.

No such defense applies here: the only mathematical knowledge needed to comprehend that Marshallian atomism is incompatible with a horizontal demand curve for the firm is elementary calculus.

The responses I have received on this point from neoclassical economists to date have been disingenuous. At best, they have referred to Stigler's attempt to recast perfect competition as the limiting case as the number of firms in an industry increases (discussed in

the next section).[1] At worst, they have claimed that the laws of mathematics do not apply to economics.[2]

The latter claim is of course nonsense for an approach to economics which, from its founding father to today's leading exponents, exalted itself over its rivals because it *was* mathematical:

those economists who do not know any mathematics … can never prevent the theory of the determination of prices under free competition from becoming a mathematical theory. Hence, they will always have to face the alternative either of steering clear of this discipline … or of tackling the problems of pure economics without the necessary equipment, thus producing not only very bad pure economics but also very bad mathematics. (Walras 1900 [1954]: 47)

This raises the question of why neoclassical economists defend commencing an education in economics with such bad mathematics? We expect it is because the fantasy of perfect competition is essential to fulfilling the vision of rational self-interested behavior being compatible with welfare maximization. If one admits that the individual firm faces a downward-sloping demand curve, then the elimination of deadweight loss that is the hallmark of perfect competition can't possibly be compatible with individual profit-maximization.

This is easily illustrated using another standard mathematical technique, the Taylor series expansion.[3]

Perfect competitors aren't profit maximizers

Consider a competitive industry where all firms are producing at the "perfect competition" level where price equals marginal cost. In general, profit for the i^{th} firm is:

$$\pi_i(q_i) = P(Q)\cdot q_i - TC(q_i) \quad (0.2)$$

What happens to the i^{th} firm's profits if it changes its output by a small amount δq_i? Under the Marshallian condition of atomism, industry output also changes by the same amount. The change in profit $\delta\pi(\delta q_i)$ is thus

$$\pi_i(q_i+\delta q_i) - \pi_i(q_i) = \left(P(Q+\delta q_i)\cdot(q_i+\delta q_i) - TC(q_i+\delta q_i)\right) - \left(P(Q)\cdot q_i - TC(q_i)\right) \quad (0.3)$$

This can be approximated by applying the first order Taylor series expansion, and by making the substitution that, at this output level, price equals marginal cost: $P(Q) = TC'(q_i)$. The symbolic mathematics engine in Mathcad makes fast work of this approximation:[4]

[1] A referee for the *Economic Journal* wrote that "we always consider the perfect competition case as a polar case which represents an extreme scenario, and is largely a benchmark. I would prefer to see the equation: (AR - MR)/AR = 1 /(nE), so that for E at a normal value of say 2, and n at 1000, then the divergence of AR and MR is $1/20^{t}$ of 1%. Then price equals MR seems a pretty good approximation!"

[2] A referee for the *Journal of Economic Education* commented that "Stigler's many attempts to save neoclassical theory have always caused more problems than they have solved. His version of the chain rule is contrary to the partial equilibrium method and thus is irrelevant".

[3] This proof was first developed by John Legge, of La Trobe University.

$$\left[P(Q+\delta q_i)\cdot(q_i+\delta q_i)-TC(q_i+\delta q_i)\right]-\left(P(Q)\cdot q_i-TC(q_i)\right)\ \Big|\ \begin{array}{l}\text{expand}\\ \text{series},\delta q_i,1\\ \text{substitute},\frac{d}{dq_i}TC(q_i)=P(Q)\end{array}\ \rightarrow q_i\cdot\delta q_i\cdot\frac{d}{dQ}P(Q)$$

Figure 2: Mathcad's symbolic solution for change in a firm's profit from perfect competition output level

Therefore $\delta\pi(\delta q_i)\approx q_i\cdot\delta q_i\cdot\frac{d}{dQ}P$. Since $\frac{d}{dQ}P<0$, if $\delta q_i<0$—if, in words, the firm reduces its output—its profit will rise. Thus the output level at which price equals marginal cost is not a profit maximum for the individual competitive firm, and if such a firm is indeed a profit maximizer, it will reduce its output below this level.

Some neoclassical economists have thrown the "perfect knowledge" assumption at us at this point: perfectly informed consumers will instantly stop buying from the firm that has reduced its output and increased its price, and switch to those that are still setting price equal to marginal cost. But this argument is still based on the "horizontal demand curve" assumption, which itself is a furphy,[5] and the market price in the model has already risen because of the change in output by one firm—there is no "cheaper supplier" to whom omniscient consumers can turn.

"Price equals marginal cost" is, therefore, not an equilibrium under the Marshallian assumption of atomism. As a result, the coincidence of collective welfare and the pursuit of individual profit is impossible: if neoclassical economists want to pull that particular rabbit out of a hat, they need another hat. Stigler attempted to provide one.

Stigler's limiting case

Stigler, of course, was not trying to bury perfect competition when he showed that $\frac{dP}{dq_i}=\frac{dP}{dQ}$: he was one of the pre-eminent defenders of the neoclassical model against empirically-oriented researchers like Eiteman and Means (see Freedman 1998). He therefore devised an alternative explanation of perfect competition, as the limiting case of competition as the number of firms in an industry increased. His analysis, shown in **Figure 1**, footnoted the derivation of the expression shown in **Figure 3**:

$$\text{Marginal revenue} = \text{Price} + \frac{\text{Price}}{\text{Number of sellers} \times \text{Market elasticity}}$$

Figure 3: Stigler's expression for marginal revenue (Stigler 1957: 8)

[4] We are using a symbolic mathematics program both to reduce the need for some tedious manual calculations, and because on several occasions, neoclassical economists have disputed the results of manual calculations—by in effect distorting the definition of a derivative.

[5] "furphy" is a delightful Australian word meaning "an irrelevant or minor issue raised to specifically divert attention away from the real issue". It deserves wider currency—especially amongst economists!

Stigler then asserted that "this last term goes to zero as the number of sellers increases indefinitely" (Stigler 1957: 8). Marginal revenue for the i^{th} firm thus converges to market price. Perfect competition thus appeared to be saved, despite a downward-sloping firm's demand curve: profit-maximizers would set marginal cost equal to marginal revenue, and this would converge to price as more firms entered a market.

Stigler's convergence argument is technically correct, but in conflict with the proof shown above that "price equals marginal cost" is *not* a profit maximum for the individual firm. The resolution of this conflict led to Keen's first truly original contribution to this literature: *the proposition that equating marginal revenue and marginal cost maximizes profit is also a furphy*.

Equating MC and MR doesn't maximize profits

Generations of economists have been taught the simple mantra that "profit is maximized by equating marginal cost and marginal revenue". The proof simply differentiates (0.2) with respect to q_i. However, the individual firm's profit is a function, not only of its own output, but of that of all other firms in the industry. This is true regardless of whether the firm reacts strategically to what other firms do, and regardless of whether it can control what other firms do. The objectively true profit maximum is therefore given by the zero of the *total* differential: the differential of the firm's profit with respect to total industry output.

We stress that this issue is independent of whether the individual firm can or cannot work out this maximum for itself, whether the firm does or does not interact with its competitors, and whether the firm does or does not control the variables that determine the profit maximum. Given a mathematically specified market inverse demand function that is a function of the aggregate quantity supplied to the market, and a mathematically specified total cost function for the individual firm that is a function of its output, the question "what is the level of the firm's output that maximizes its profit?" is completely independent of the question of "will the firm, in any given environment, or following any given behavioral rule, actually determine or achieve this level?". That objective, profit-maximizing level is given by the zero of the *total* differential of profit:

$$\frac{d}{dQ}\pi\left(q_i\right)=\frac{d}{dQ}\left(P\left(Q\right)q_i-TC\left(q_i\right)\right)=0 \tag{0.4}$$

This total derivative is the sum of n partial derivatives in an n-firm industry:

$$\frac{d}{dQ}\pi\left(q_i\right)=\sum_{j=1}^{n}\left\{\left(\frac{\partial}{\partial q_j}\pi\left(q_i\right)\right)\cdot\frac{d}{dQ}q_j\right\} \tag{0.5}$$

In the Marshallian case, atomism lets us set $\frac{d}{dQ}q_j=1\ \forall \mathrm{j}$ (we address the Cournot case in section 0). Expanding (0.5) yields

$$\frac{d}{dQ}\pi\left(q_i\right)=\sum_{j=1}^{n}\left(\frac{\partial}{\partial q_j}\left(P\left(Q\right)q_i-TC\left(q_i\right)\right)\right) \tag{0.6}$$

Continuing with the product rule, (0.6) can be expanded to:

$$\frac{d}{dQ}\pi(q_i)=\sum_{j=1}^{n}\left(P(Q)\frac{\partial}{\partial q_j}q_i+q_i\cdot\frac{\partial}{\partial q_j}P(Q)-\frac{\partial}{\partial q_j}TC(q_i)\right) \quad (0.7)$$

Under the Marshallian assumption of atomism, the first term in the summation in (0.7) is zero where $j\neq i$, and $P(Q)$ where $j=i$. The second term is equal to $q_i\cdot\frac{d}{dQ}P(Q)\ \forall j$; the third is zero where $j\neq i$, and equal to $\frac{d}{dq_i}TC(q_i)$ (or marginal cost $MC(q_i)$) where $j=i$. Equation (0.7) thus reduces to

$$\frac{d}{dQ}\pi(q_i)=P(Q)+n\cdot q_i\cdot\frac{d}{dQ}P(Q)-MC(q_i) \quad (0.8)$$

The true profit maximum—under the Marshallian condition of atomism—is thus given by equation (0.9):

$$\pi(q_i)_{\max}:MC(q_i)=P+n\cdot q_i\cdot\frac{dP}{dQ} \quad (0.9)$$

The error in the standard "Marshallian" formula is now obvious: it omits the number of firms in the industry from the expression for the individual firm's marginal revenue. With this error corrected, the correct profit-maximizing rule for a competitive firm is very similar to that for a monopoly: set marginal cost equal to *industry level* marginal revenue.[6]

Monopoly, competition, profit and hyper-rationality

Neoclassical economics assumes that, given revenue and cost functions, there is some output level that will maximize profits, and another that will maximize social welfare (by eliminating deadweight loss).[7] The argument that the two coincide under perfect competition has been shown to be nonsense. So too is the argument that a single rational firm could work out the profit maximum, but a bunch of rational *non-interacting* firms couldn't, as the calculus in the previous section shows.

Of course, an objection can be made to the above mathematical logic that solving equation (0.9) requires knowledge of the number of firms in the industry, which the individual competitive firm can't be assumed to have.[8] Here, we can turn Milton Friedman's methodological defense of the theory of the firm against itself. Friedman, as is well known, argued that while firms didn't in fact do calculus to work out their profit-maximizing output

[6] Though not necessarily identical, since $n\cdot q_i$ only equals Q if $q_i=Q/n\ \forall \mathrm{i}$. This impact of dispersal in firm size may explain some of the simulation results shown later.

[7] We use standard undergraduate terms here because the analysis we are challenging is, up to this point, that served up to undergraduates. We address game theoretic concepts later.

[8] Equation (1.9) can be put in another form which partly addresses this criticism, and also emphasizes the error in the conventional formula. The profit-maximizing level of output is not to equate firm-level marginal revenue and marginal cost, but to make the gap between them a fraction of the gap between price and marginal cost: $MR(q_i)-MC(q_i)=\frac{n-1}{n}(P-MC)$. The fraction tends to 1 as $n\to\infty$, so the more "competitive" an industry is, the easier it is to apply this formula.

levels, we could model their behavior "as if" they did, because unless the behavior of businessmen in some way or other approximated behavior consistent with the maximization of returns, it seems unlikely that they would remain in business for long. (Friedman 1953: 22)

We are not arguing that firms do the calculus to work out this profit-maximizing level either.[9] Instead, we are simply showing that the calculus *can* be done, and the profit-maximizing level is not the one asserted by neoclassical economists. However, it is possible now—in a way that wasn't possible in 1953—to actually carry out Friedman's "billiard players" experiment. Citing him again:

Now, of course, businessmen do not actually and literally solve the system of simultaneous equations in terms of which the mathematical economist finds it convenient to express this hypothesis, any more than leaves or billiard players explicitly go through complicated mathematical calculations or falling bodies decide to create a vacuum. The billiard player, if asked how he decides where to hit the ball, may say that he "just figures it out" but then also rubs a rabbit's foot just to make sure; and the businessman may well say that he prices at average cost, with of course some minor deviations when the market makes it necessary. The one statement is about as helpful as the other, and neither is a relevant test of the associated hypothesis. (Friedman 1953: 22)

A "relevant test of the associated hypothesis" is to set up a virtual market that conforms to neoclassical assumptions—with a static downward sloping market demand curve, and given cost functions subject to diminishing marginal productivity, so that there is indeed a profit-maximizing level of output for each firm—and see what happens. **Figure 4** shows a Mathcad program that implements this.[10]

[9] In fact, we argue later that the assumption that there is some profit-maximizing level for a firm is a furphy. The profit-maximizing strategy for actual firms is simply sell as much as possible, at the expense where possibl

[10] The behavior modeled was deliberately made as simple as possible, to avoid the rejoinder that the results were the product of our algorithm rather than raw profit-motivated behavior. It could only have been simpler by having each firm vary its output by one unit at each time step—a modification which, as it happens, results in a much slower but absolute convergence to the Keen equilibrium.

$$
\begin{array}{l}
\text{Firms} := \left| \begin{array}{l}
\text{Seed}(\text{rand}) \\
\text{for } i \in \text{firms}_{\text{min}}, \text{firms}_{\text{min}} + \text{firms}_{\text{steps}} .. \text{firms}_{\text{max}} \\
\quad \left| \begin{array}{l}
Q_0 \leftarrow \left| \begin{array}{ll} \text{round}\left(\text{runif}\left(i, q_K(i), q_C(i)\right)\right) & \text{if } i > 1 \\ q_C(i) & \text{otherwise} \end{array} \right. \\
p_0 \leftarrow \left| \begin{array}{ll} P\left(\sum Q_0, a, b\right) & \text{if } i > 1 \\ P\left(q_C(i), a, b\right) & \text{otherwise} \end{array} \right. \\
dq \leftarrow \left| \begin{array}{ll} \text{round}\left(\text{rnorm}\left(i, 0, \dfrac{q_C(i)}{100}\right)\right) & \text{if } i > 1 \\ \dfrac{q_C(i)}{100} & \text{otherwise} \end{array} \right. \\
\text{for } j \in 0.. \text{runs} - 1 \\
\quad \left| \begin{array}{l}
Q_{j+1} \leftarrow Q_j + dq \\
p_{j+1} \leftarrow \left| \begin{array}{ll} P\left(\sum Q_{j+1}, a, b\right) & \text{if } i > 1 \\ P\left(Q_{j+1}, a, b\right) & \text{otherwise} \end{array} \right. \\
dq \leftarrow \overrightarrow{\left[\text{sign}\left[\left(p_{j+1} \cdot Q_{j+1} - p_j \cdot Q_j\right) - \left(\text{tc}\left(Q_{j+1}, i\right) - \text{tc}\left(Q_j, i\right)\right)\right] \cdot dq\right]} \\
F_{j, i-1} \leftarrow Q_j
\end{array} \right.
\end{array} \right. \\
F
\end{array} \right.
\end{array}
$$

Figure 4: Simulation of instrumental profit maximizers

Working through the program line by line:

1. A random number generator is seeded
2. A *for* loop iterates from a minimum number to a maximum number of firms
3. If there is more than one firm in the industry, each firm is randomly allocated an initial output level. The amounts are uniformly distributed from a minimum of the Keen prediction for a profit-maximizing firm, q_K to a maximum of the neoclassical prediction q_C.
4. If there is only one firm in the industry, its output starts at the level predicted by the neoclassical model—which coincides with q_K.
5. An initial market price is set, based on the sum of initial outputs.
6. Line 6 sets the market price in the case of a monopoly.
7. Each firm is randomly allocated an amount by which it varies output. The distribution has a mean of zero and a standard deviation of 1% of the neoclassical prediction for a profit-maximizing firm's output (this is the last aspect of the program that involves probability).
8. Line 8 allocates a change amount of 1% of the predicted output for a monopoly.

9. A *for* loop iterates over a number of runs where each firm varies its output trying to increase its profit from the initial level.
10. Firstly each firm adds its change amount to its initial output. This is a vector operation: if there are 100 firms in the industry Q_0 is a vector with 100 initial output amounts, and *dq* is a vector with 100 (positive or negative) output changes.
11. A new market price is calculated on the basis of the new aggregate output level.
12. Line 12 again allows for a monopoly.
13. Each firm then calculates whether its profit has risen or fallen as a result of its change in output, and the collective impact of all the changes in output on the market price. If a firm finds that its profit has risen, it continues to change output in the same direction; if its profit has fall, it changes its output by the same amount but in the opposite direction.
14. Each step in the iteration is stored in a multi-dimensional array.[11]
15. The multidimensional array is returned by the program.

The program was run with identical cost functions for each firm, set up so that the market aggregate marginal cost curve was independent of the number of firms in the industry (we return to this issue in the Appendix). The number of firms was varied from 1 to 100. The eventual aggregate output at the end of 1000 iterations is shown in **Figure 5**, and the corresponding market price is shown in **Figure 6**, against the predictions of the Neoclassical and the Keen approach respectively.

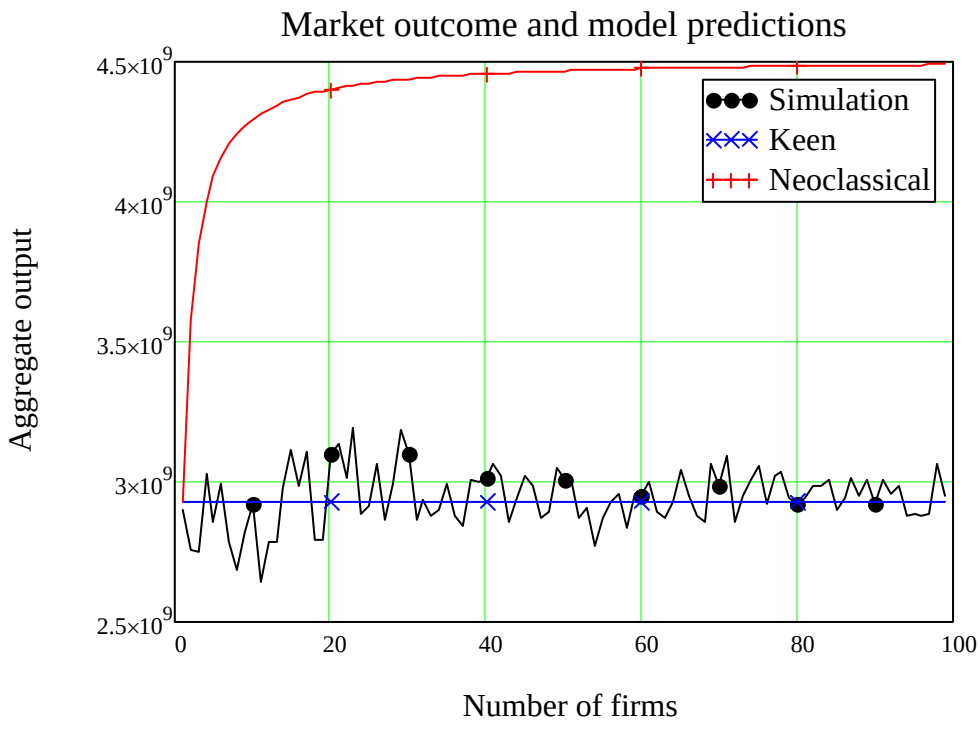


Figure 5: Aggregate output

[11] In effect, F is a matrix where the j[th] and i[th] column contains the vector of outputs by an i firms industry at the j[th] iteration.

As is obvious, the number of firms in an industry had no impact on the eventual market output level or price: the Neoclassical prediction that price would converge to the level at which price equals marginal cost clearly was not fulfilled.

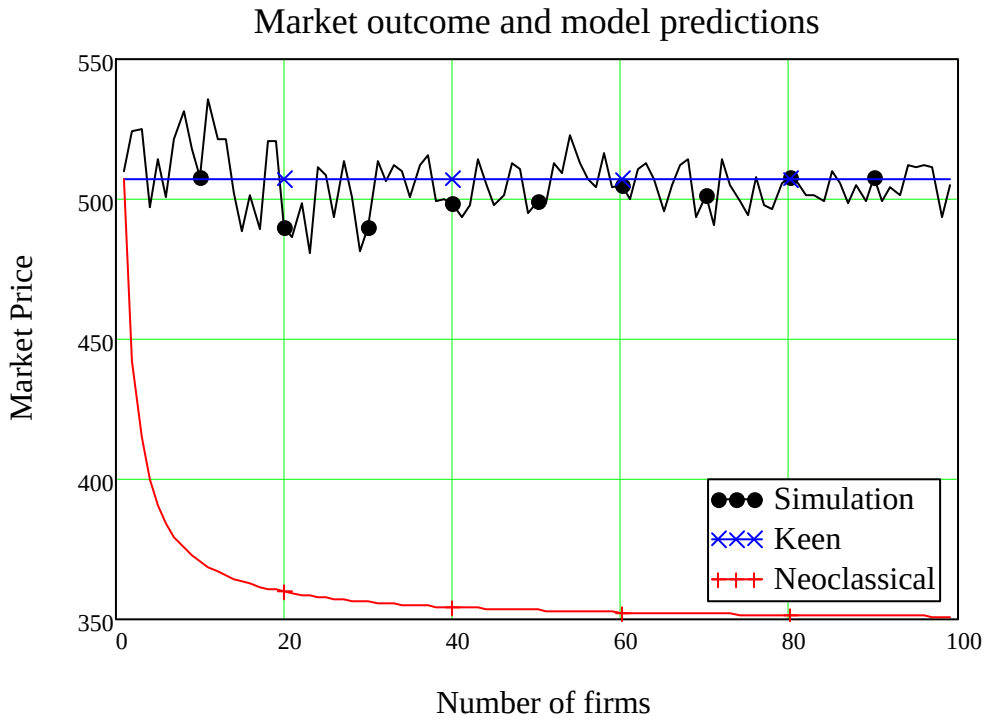


Figure 6: Market price

Some neoclassical referees thought that the results occurred because, though all firms were acting independently, they were all doing the same thing (reducing output from the initial level), and thus acting in a semi-collusive way.[12] In fact, as **Figure 7** and **Figure 8** show, though the average outcome conformed to Keen's predictions, the individual firms all pursued very different strategies. The aggregate outcome, which contradicted the neoclassical prediction and confirmed Keen's, was the result of quite diverse individual firm behavior—despite all firms having identical cost functions.

Figure 7 shows the output levels of 3 randomly chosen firms from the 100 firm industry, the average for all firms, and the predictions of the Keen and neoclassical formulae. Firm 1 began near the neoclassical output level, rapidly reduced output towards the "Keen" level, but then reversed direction; Firm 2 began halfway between the neoclassical and Keen predictions, then reduced output below the Keen level and stayed there; Firm 3 began closer to the neoclassical level and meandered closer to the Keen level.

[12] A referee for the *Economic Journal* commented that "if firms act the same way, they will all get higher profits if and only if they reduce outputs. Then the algorithm will continue to lead them to the monopoly outcome since there is no chance any firm can realize the true impact of its own output change. Thus the result is not surprising."

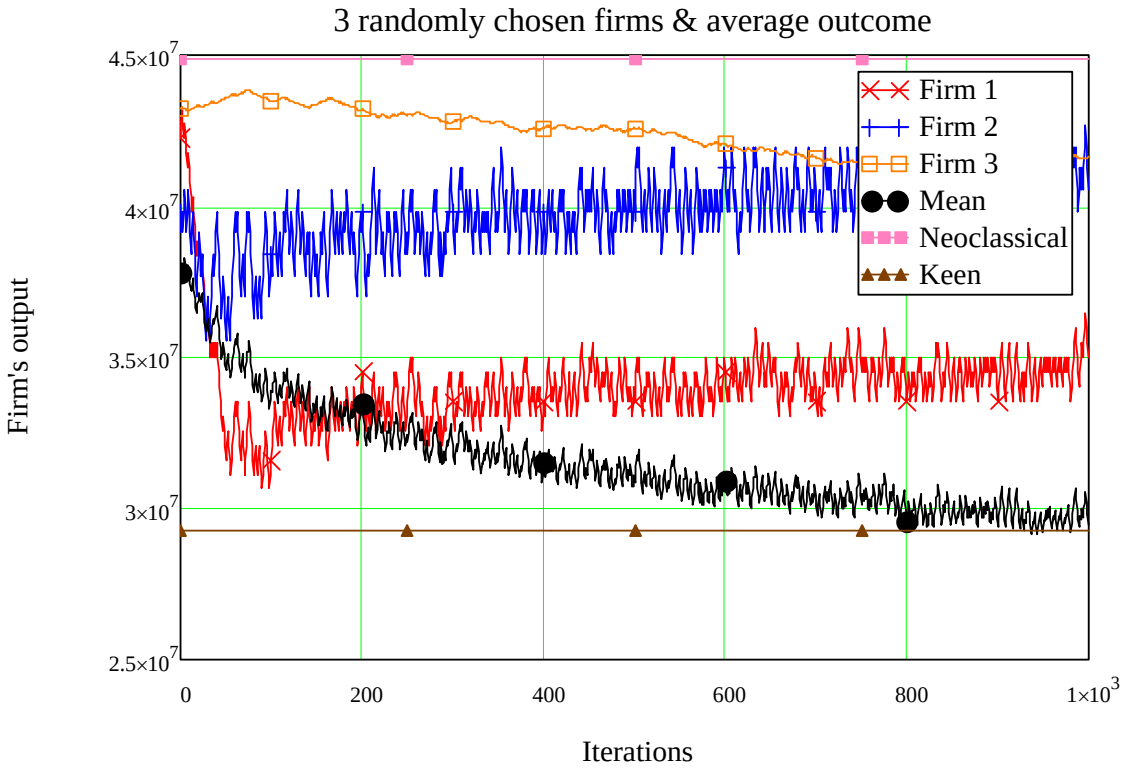


Figure 7: Firm outputs in 100 firm industry

The sole source of the volatility of each firm's behavior is the complex impact of interactions between firms, in the context of a very simply defined market—there is no random number generator causing this volatility. As **Figure 8** shows, each firm made its changes in response to the impact of both its changes in output, and the collective changes in output, on its profit. Some firms made larger profits than others—notably the firms with the larger output made the larger profits. However, the average profit was much higher than predicted by the neoclassical model.

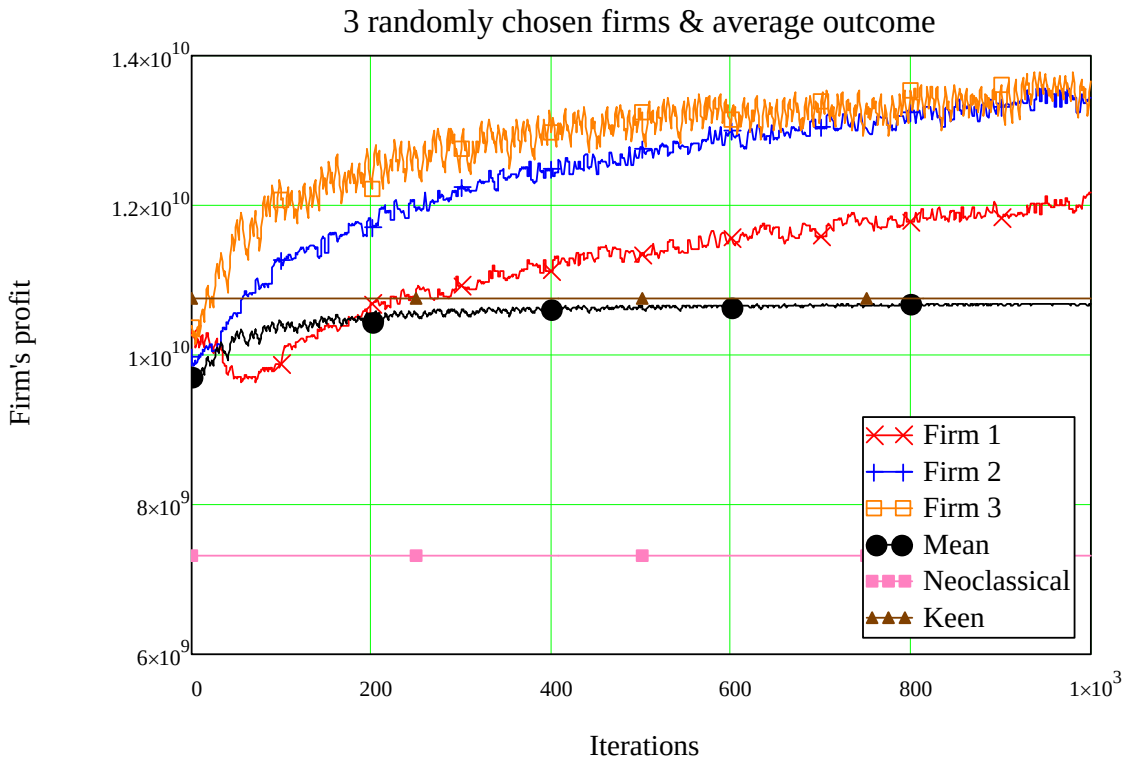


Figure 8: Firm profits in 100 firm industry

This model indicates that, in this game of competitive profit maximization, the virtual equivalents of Friedman's "billiard players" follow the laws of mathematics in their search for profits, as Friedman argued. However, these laws of mathematics are incompatible with the beliefs of neoclassical economists.

Since hyper-rational profit-maximizers cannot be relied upon to save neoclassical belief, there are only two avenues left: irrational behavior, and Cournot-Nash game theory.

Price-taking behavior is irrational

A regular neoclassical rejoinder to our analysis has been that we are "cheating" by not assuming rational, price-taking behavior. Our standard reply that the assumption of "price-taking" behavior is itself cheating, with regard to the laws of mathematics: as shown in Section 0, the assumption that $P`(q_i) = 0$ is incompatible with the assumption of a downward-sloping market demand curve ($P`(Q) < 0$). However, it is also easily shown that "price-taking behavior" is irrational.

The assumption of price-taking behavior appears regularly in neoclassical economics, from the level of Marshallian analysis through to the foundations of general equilibrium analysis (see for example Mas-Colell et al 1995: 314, 383). Neoclassical economists do not seem to realize that this is a classic "rabbit in the hat" assumption: if it is assumed, then the "perfectly competitive" result of price equaling marginal cost follows, regardless of how many firms there are in the industry.

The essence of price-taking is the belief that a firm's change in its output doesn't affect market price: this amounts to setting $\frac{\partial}{\partial q_j} P(Q) = 0$ in equation (0.7). This results in the "profit-maximizing strategy" of setting price equal to marginal cost, independently of the number of firms—that is, once this assumption is made, *even a monopoly produces where price equals marginal cost*. This behavior is clearly irrational for a monopoly, and it is only the "fog of large numbers"—the confusion of infinitesimals with zero, as Keen noted in *Debunking Economics*—that led neoclassical economists to regard price-taking as rational behavior for competitive firms.

Figure 9 illustrates that price-taking behavior is *irrational*: an agent who behaves this way is necessarily making a logical error. If the market demand curve slopes downwards, then the *a priori* rational belief is that *any* increase in output by the firm will depress market price.

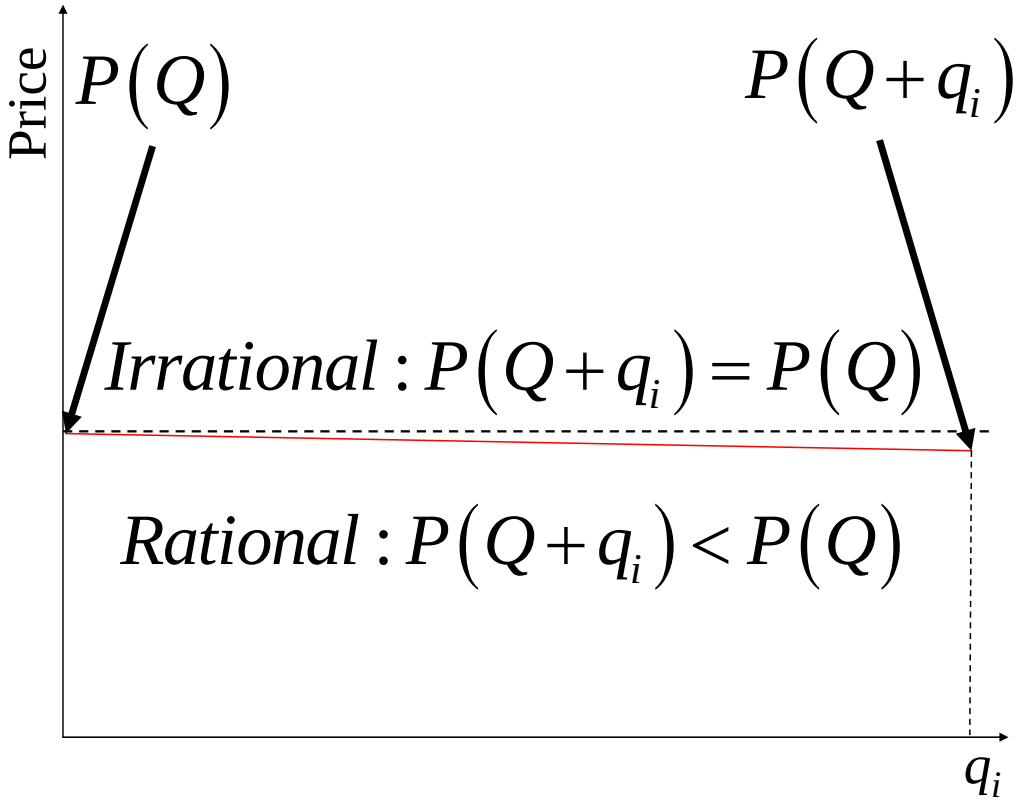


Figure 9: Irrationality of "price-taking" behavior

The desired neoclassical of price equal to marginal cost is thus dependent on irrational behavior (in the context of Marshallian competition—we address Cournot competition later). We quantify the degree of irrationality needed by modifying the program shown in **Figure 4**, so that a proportion of firms actually do behave irrationally: if a strategy has caused an increase in profit, a fraction of firms respond by *reversing* that strategy.

The modified program is shown in **Figure 10**. The outer loop (line 2) now iterates the counter *i* from 0 to 50, with the value representing the fraction of firms who behave irrationally

at each iteration. The only change to the inner loop is that the change in output by each firm is now reversed for *i*% of firms at each iteration.[13]

```
Firms:= | Seed(rand)
        | for i ∈ 0..50
        |   Q_0 ← round(runif(firms, q_K(firms), q_C(firms)))
        |   p_0 ← P(∑Q_0, a, b)
        |   dq ← round(rnorm(firms, 0, q_C(firms)/100))
        |   for j ∈ 0..runs − 1
        |     Q_{j+1} ← Q_j + dq
        |     p_{j+1} ← P(∑Q_{j+1}, a, b)
        |     dq ← [sign[runif(firms, −i/100, −i/100 + 1)·[(p_{j+1}·Q_{j+1} − p_j·Q_j) − (tc(Q_{j+1}, firms) − tc(Q_j, firms))]]·dq
        |     F_{j,i} ← Q_j
        | F
```

Figure 10: Analyzing the impact of irrationality

Figure 11 shows the aggregate outcome for a 100 firm industry. With no irrationality, the industry produces the amount predicted by the Keen formula. Output then increases almost monotonically as the degree of irrationality rises—until, when 20 per cent of firms are behaving irrationally at each iteration, market output converges to near the neoclassical output level.

For a degree of irrationality between 20% and 45%, the neoclassical outcome continues to dominate the simulation results. Then as irrationality rises above this level, the market effectively follows a random walk—where, curiously, profits in general tend to be *higher* than what would apply if each firm equated marginal revenue and marginal cost.

[13] The function call $runif\left(firms, -i/100, -i/100+1\right)$ generates a vector of numbers between $-i/100$ and $1-i/100$; when *i=0*, all these numbers will be positive and thus not affect the value of the *sign()* function; when *i>0*, *i*% of these numbers will be negative and thus the sign of the *sign()* function will be reversed. The firms that have this randomly assigned negative number against their output change will increase output at the next step if profit rose when the decreased output on the previous step (and vice versa). This is instrumentally irrational behavior.

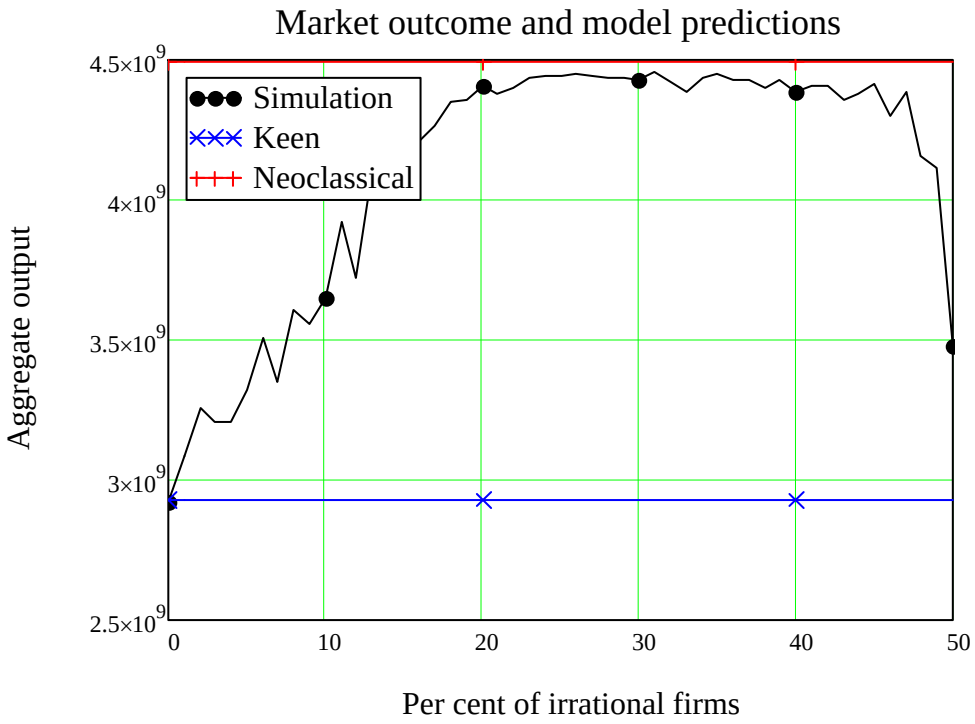


Figure 11: Market output as a function of the degree of irrationality

Figure 12 shows the behavior of three randomly chosen firms, and the average behavior, at a 20% level of irrationality—i.e., when one firm in five reverses any strategy that benefited it on the previous iteration.

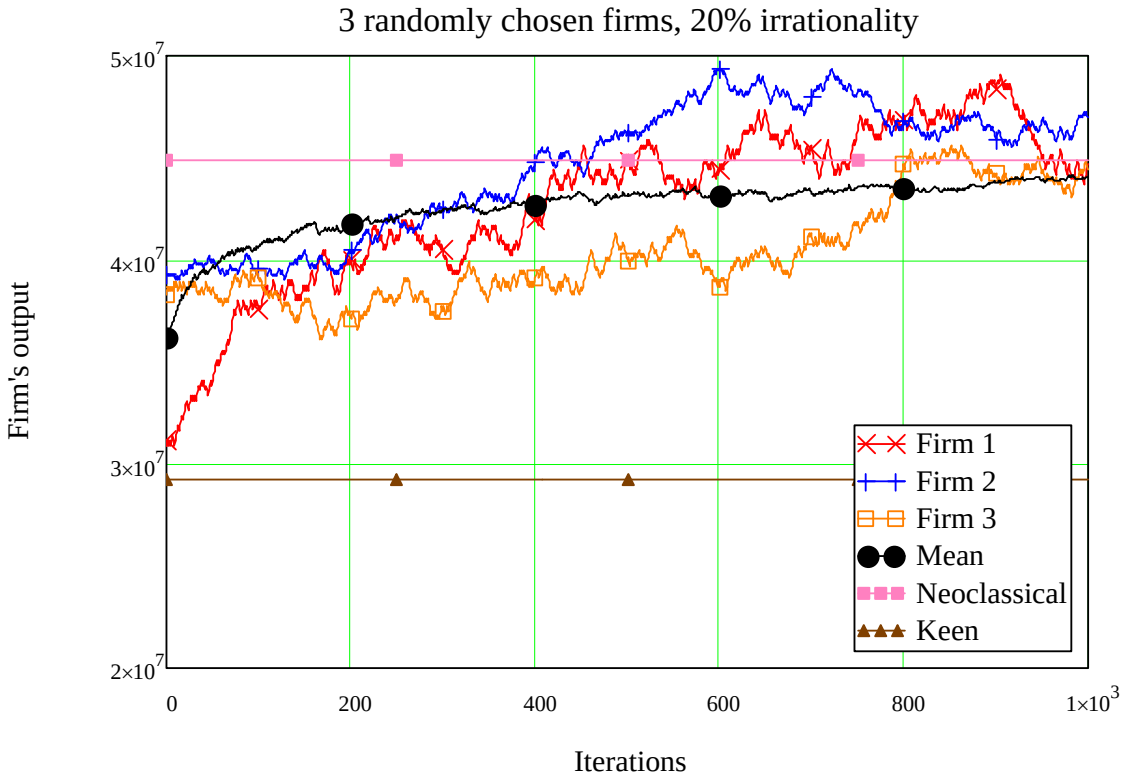


Figure 12: Sample outputs at 20% irrationality

Figure **13** shows the impact that a degree of irrationality of 20% has on firms' profits. Profit falls throughout the run, until by the end, it is almost (but not quite) as low as that caused by equating marginal revenue and marginal cost.

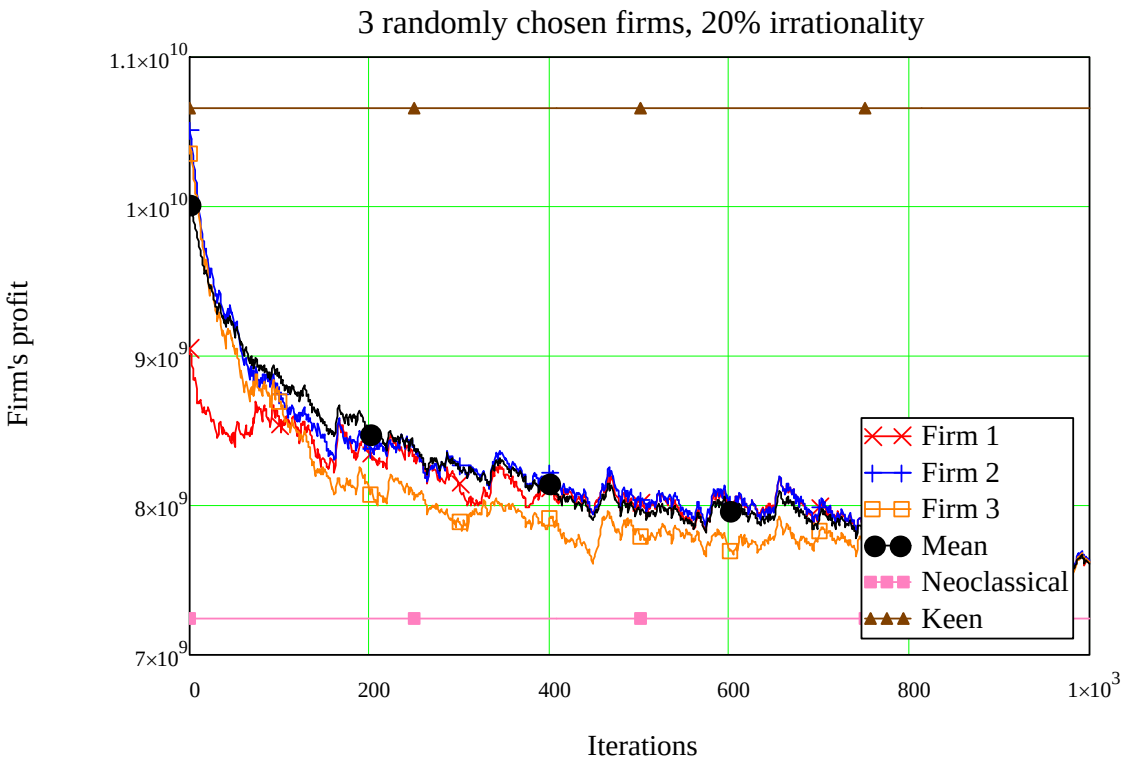


Figure 13: Impact of 20% irrationality on firms' profits

Ironically, higher profits apply if firms simply follow a random walk than if they apply the neoclassical formula (see **Figure 14**).

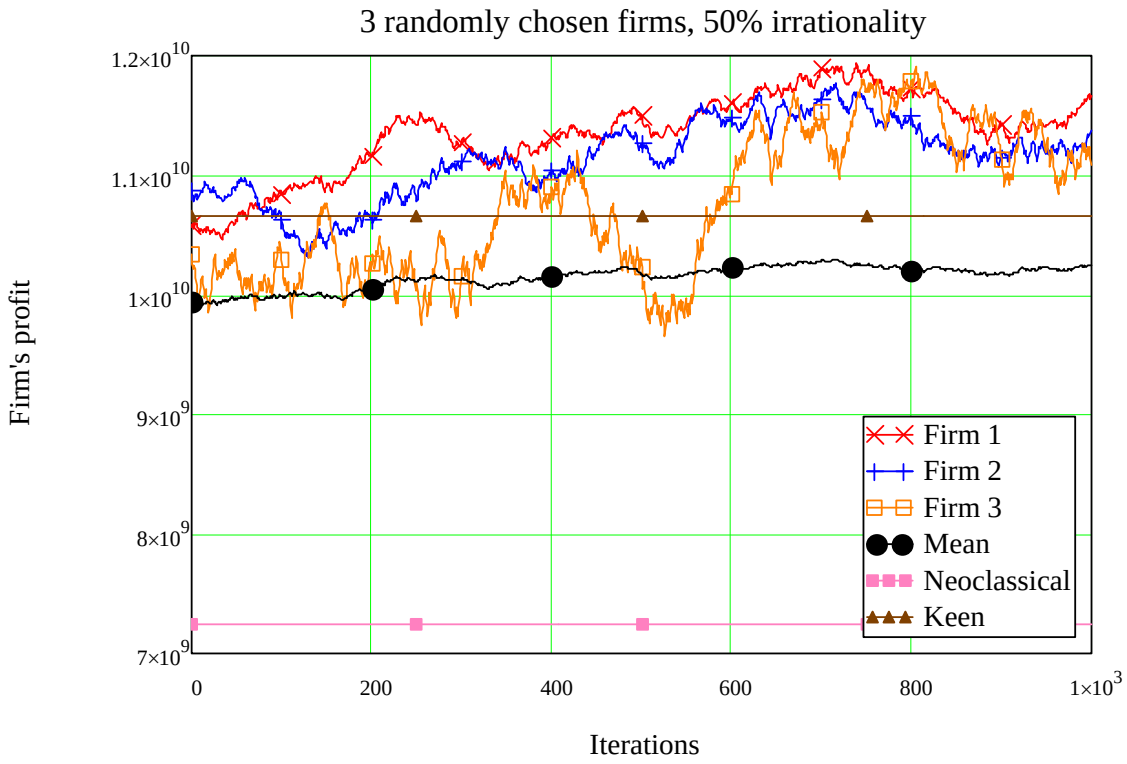


Figure 14: Firm profits with 50% irrationality

A degree of irrational behavior thus saves the neoclassical preferred outcome of price equal to marginal cost—though with some collateral damage, since it is now clearly neither profit-maximizing, nor rational. The question remains, what might help ensure this level of irrationality? Cournot-Nash game theory appears to provide an answer in strategic interactions between firms—though this answer is only unequivocal at a very superficial level of analysis.

Strategic interaction and competition

Unlike the strictly false Marshallian model of competition, Cournot-Nash game theory provides a prima facie sound basis for "perfect competition" as the outcome of strategic interactions between competitors. In Cournot-Nash game theoretic analysis, firms decide their own actions on the basis of the expected reactions of other firms, in such a way that each firm's best response is to set $MR(q_i) = MC(q_i)$. This is profit-maximizing for the firm, *in the context of the expected response of competitors to its actions*, though it results in a lower level of profit than if firms "collude" to share the monopoly level of output between them.

Up to this point, our contribution has been to show that what neoclassicals call "collusive behavior" can actually result from firms *not* reacting strategically to what other firms do—in the notation of the early part of this paper, when firms set $\frac{\partial q_i}{\partial q_j} = 0 \ \forall i \neq j$

This paradox—that what neoclassical theory labels "collusion" actually occurs when firms do not react to each other—inspired us to attempt to integrate (corrected) Marshallian and Cournot-Nash theory, by making the level of strategic interaction between firms a variable. Defining the response of the i[th] firm to an output change by the j[th] firm as $\theta_{i,j} = \frac{\partial q_i}{\partial q_j}$,

we then had to rework the expression for profit into one that depended entirely upon the level of strategic interaction.[14] The result of this research was a second original contribution, a generalized formula for profits in terms of the level of strategic interaction—and the discovery that the optimal level of interaction, in the context of identical firms, was zero. The derivations involved are quite complex, and they are reproduced below in their entirety.

We start from the same position as equation (0.4). For profit-maximization, we require the zero of $\frac{d}{dQ}\pi(q_i)$. We then expand this as per equation (0.5), but rather than then setting $\frac{d}{dQ}q_j = 1 \ \forall \mathrm{j}$, we work out what $\frac{d}{dQ}q_j$ is in terms of the strategic reaction coefficient $\theta_{i,j}$:

$$\frac{d}{dQ}q_i = \sum_{j=1}^{n}\frac{\partial}{\partial q_j}q_i$$
$$= \sum_{j=1}^{n}\theta_{i,j} \qquad (0.10)$$

As a result, our next equation differs from equation (0.6):

$$\frac{d}{dQ}\pi(q_i) = \sum_{j=1}^{n}\left(\frac{\partial}{\partial q_j}\left(P(Q)\cdot q_i - TC(q_i)\right)\cdot\frac{d}{dQ}q_j\right)$$
$$= \sum_{j=1}^{n}\left(\frac{\partial}{\partial q_j}\left(P(Q)\cdot q_i\right)\cdot\frac{d}{dQ}q_j\right) - \sum_{j=1}^{n}\left(\frac{\partial}{\partial q_j}TC(q_i)\cdot\frac{d}{dQ}q_j\right)$$

(0.11)

Working firstly with the total cost component, $\frac{\partial}{\partial q_j}TC(q_i) = 0 \ \forall i \neq j$ and $\frac{\partial}{\partial q_j}TC(q_i) = MC(q_i) \ \forall i = j$. Thus the cost component of the profit formula reduces to:

$$\sum_{j=1}^{n}\left(\frac{\partial}{\partial q_j}TC(q_i)\cdot\frac{d}{dQ}q_j\right) = MC(q_i)\cdot\frac{d}{dQ}q_i$$
$$= MC(q_i)\cdot\sum_{j=1}^{n}\theta_{i,j} \qquad (0.12)$$

The revenue component involves some more intricate steps:

$$\sum_{j=1}^{n}\left(\frac{\partial}{\partial q_j}\left(P(Q)\cdot q_i\right)\cdot\frac{d}{dQ}q_j\right) = \sum_{j=1}^{n}\left(P(Q)\cdot\frac{\partial}{\partial q_j}(q_i)\cdot\frac{d}{dQ}q_j\right) + \sum_{j=1}^{n}\left(q_i\cdot\frac{\partial}{\partial q_j}\left(P(Q)\right)\cdot\frac{d}{dQ}q_j\right)$$

(0.13)

[14] This work was first published in Keen & Standish 2006.

$\frac{\partial}{\partial q_j}\big(P(Q)\big)$ reduces to $\frac{d}{dQ}\big(P(Q)\big)$ as before, though the logic is slightly more complicated:

$$\begin{aligned}\frac{\partial}{\partial q_j}\big(P(Q)\big) &= \frac{\partial}{\partial Q}\big(P(Q)\big)\cdot\frac{\partial}{\partial q_j}Q \\ &= \frac{\partial}{\partial Q}\big(P(Q)\big)\cdot 1 \\ &= \frac{d}{dQ}P(Q)\end{aligned} \qquad (0.14)$$

Making this substitution into (0.13), and using P rather than P(Q) for the sake of clarity, yields:

$$\begin{aligned}&\sum_{j=1}^{n}\left(P\cdot\frac{\partial}{\partial q_j}(q_i)\cdot\frac{d}{dQ}q_j\right)+\sum_{j=1}^{n}\left(q_i\cdot\frac{dP}{dQ}\cdot\frac{d}{dQ}q_j\right) \\ &= P\cdot\sum_{j=1}^{n}\left(\theta_{i,j}\cdot\frac{d}{dQ}q_j\right)+q_i\cdot\frac{dP}{dQ}\cdot\sum_{j=1}^{n}\left(\frac{d}{dQ}q_j\right)\end{aligned} \qquad (0.15)$$

Care has to be taken with expanding the expression $\frac{d}{dQ}q_j$ in (0.15), since $\frac{d}{dQ}q_j=\sum_{j=1}^{n}\theta_{j,i}$, but the *i* suffix here is just a placeholder for iterating over the *n* firms in the industry. We therefore make the substitution of *k* for *i* in this subscript so that we define $\frac{d}{dQ}q_j=\sum_{k=1}^{n}\frac{\partial}{\partial q_k}q_j=\sum_{k=1}^{n}\theta_{j,k}$.

$$\begin{aligned}&P\cdot\sum_{j=1}^{n}\left(\theta_{i,j}\cdot\frac{d}{dQ}q_j\right)+q_i\cdot\frac{dP}{dQ}\cdot\sum_{j=1}^{n}\left(\frac{d}{dQ}q_j\right) \\ &= P\cdot\sum_{j=1}^{n}\left(\theta_{i,j}\cdot\sum_{k=1}^{n}\theta_{j,k}\right)+q_i\cdot\frac{dP}{dQ}\cdot\sum_{j=1}^{n}\left(\sum_{k=1}^{n}\theta_{j,k}\right)\end{aligned} \qquad (0.16)$$

Equation (0.11) finally reduces to:

$$\frac{d}{dQ}\pi(q_i)=P\cdot\sum_{j=1}^{n}\left(\sum_{k=1}^{n}\theta_{i,j}\cdot\theta_{j,k}\right)+q_i\cdot\frac{dP}{dQ}\cdot\sum_{j=1}^{n}\left(\sum_{k=1}^{n}\theta_{j,k}\right)-MC(q_i)\cdot\sum_{j=1}^{n}\theta_{i,j} \qquad (0.17)$$

The zero of this equation determines the profit maximum for any given level of strategic interaction between firms. We can now rephrase the corrected Marshallian and the Cournot-Nash profit maxima in terms of their “conjectural variation” levels.

The Marshallian substitution is rather easy. Given $\frac{\partial q_i}{\partial q_j} = 0 \ \forall i \neq j$ and $\frac{\partial q_i}{\partial q_j} = 1 \ \forall i = j$, $\sum_{j=1}^{n} \theta_{i,j} = 1$; $\sum_{j=1}^{n}\left(\sum_{k=1}^{n} \theta_{j,k}\right)$ is the trace of an identity matrix so that $\sum_{j=1}^{n}\left(\sum_{k=1}^{n} \theta_{j,k}\right) = n$; and $\theta_{i,j} \cdot \theta_{j,k} = 1 \ \forall \ i = j = k$ and zero otherwise, so that $\sum_{j=1}^{n}\left(\sum_{k=1}^{n} \theta_{i,j} \cdot \theta_{j,k}\right) = 1$. Therefore in the case of atomism, the maximum of (0.17) reduces to

$$\frac{d}{dQ}\pi\left(q_i\right) = P + q_i \cdot \frac{dP}{dQ} \cdot n - MC\left(q_i\right) = 0 \qquad (0.18)$$

This reproduces the formula derived in equation (0.9).

For the Cournot case, we start from the general situation where $\theta_{i,j} = \theta \ \forall \ i \neq j$ and $\theta_{i,i} = 1$.[15] Then $\sum_{j=1}^{n} \theta_{i,j} = 1 + \sum_{i \neq j}^{n} \theta = 1 + \left(n-1\right) \cdot \theta$; $\sum_{j=1}^{n}\left(\sum_{k=1}^{n} \theta_{j,k}\right)$ is the sum of a matrix with 1 on the diagonals and θ on the off-diagonal elements, so that $\sum_{j=1}^{n}\left(\sum_{k=1}^{n} \theta_{j,k}\right) = n + \left(n^2 - n\right) \cdot \theta = n\left(1 + \left(n-1\right) \cdot \theta\right)$. $\sum_{j=1}^{n}\left(\sum_{k=1}^{n} \theta_{i,j} \cdot \theta_{j,k}\right)$ is the sum of each column of the matrix—which is $\left(n-1\right) \cdot \theta + 1$—multiplied by each element of one of its columns, so that we have $\left(n-1\right) \cdot \theta + 1$ copies of $\left(n-1\right) \cdot \theta + 1$. Thus $\sum_{j=1}^{n}\left(\sum_{k=1}^{n} \theta_{i,j} \cdot \theta_{j,k}\right) = \left(\left(n-1\right) \cdot \theta + 1\right)^2$. Making these preliminary substitutions and factoring the common element $\left(1 + \left(n-1\right) \cdot \theta\right)$,[16] we derive:

$$\frac{d}{dQ}\pi\left(q_i\right) = \left(1 + \left(n-1\right) \cdot \theta\right) \cdot \left(P \cdot \left(1 + \left(n-1\right) \cdot \theta\right) + q_i \cdot \frac{dP}{dQ} \cdot \left(n\right) - MC\left(q_i\right)\right) \qquad (0.19)$$

Given that the Cournot-Nash "best response" results in each firm setting conventionally defined marginal revenue ($P + q_i \cdot \frac{dP}{dQ}$) equal to marginal cost, we can now

[15] The alleged neoclassical equilibrium occurs where $P = MC_i\left(q_i\right)$; for long-run equilibrium, only the most efficient scale of output applies so that marginal cost is identical for all firms, therefire all firms must produce at the same level of output $q_i = q = Q \div n$. For this to be stable, all firms must have the same level of strategic interaction with each other, $\theta_i = \theta$.

[16] Since θ lies in the range $\left[0, 1/n \cdot E\right]$, $\left(1 + \left(n-1\right) \cdot \theta\right) \neq 0$; it can therefore be factored out.

work out the corresponding value for θ. This is $\theta = \frac{1}{n \cdot E}$, where *n* is the number of firms in the industry and *E* the market elasticity of demand ($E = -\frac{P}{Q}\frac{dQ}{dP}$).

It is now also possible to work out the optimum value for θ, from the view of a profit-maximizing individual firm: what level of strategic response *should* a firm have to its rivals, given that its objective is to maximize its profit?

In this generalized case of identical firms, the answer is obvious: *the optimal value of* θ *is zero*. As shown by equation (0.18), the profit maximum is where $\frac{d}{dQ}\pi(q_i) = P + q_i \cdot \frac{dP}{dQ} \cdot n - MC(q_i) = 0$. Given equation (0.19), this is only possible for $\theta = 0$. Cournot-Nash game theory is thus "A curious game. The only winning strategy is not to play"[17]. It is therefore, on closer examination, a very poor defense of the concept of perfect competition.[18]

This interpretation is given additional weight by the observation that, though the standard "Prisoners' Dilemma" presentation implies that the Cournot strategy is stable and the Keen strategy is unstable (both in a Nash equilibrium sense), the Cournot strategy is locally unstable, while the Keen strategy is locally stable.

Local Stability and Instability

In the Cournot-Nash game-theoretic analysis of duopoly, if firms "cooperate" and split the monopoly-level output, they make equally high profits. However, each firm has an incentive to "defect" and produce a larger amount where its marginal revenue equals its marginal cost, because it will make a higher profit still—if the other firm continues to produce its share according to the monopoly formula. This gives both firms an incentive to defect, resulting in both producing where marginal revenue equals marginal cost. This results in a lower profit for each firm than when they split the monopoly output between them, but it is a globally stable strategy, whereas all other strategy combinations are unstable.

As a result, output is higher and price lower under duopoly than monopoly, and the limit of this process as the number of firms increases is "perfect competition". This is illustrated with the example of a duopoly facing a linear market demand curve $P(Q) = a - b \cdot Q$, with identical quadratic total cost functions $tc(q) = k + c \cdot q + \frac{1}{2} d \cdot q^2$. Figure 9 shows the output combinations produced by two firms producing at either the Cournot or Keen predicted level, in terms of the demand and cost arguments.

[17] For those who do not know, this is a line from the 1980s movie *War Games*.

[18] It may be thought that this result is an artifact of an accidental aggregation effect from using the same reaction coefficient for all firms; we refute this by generalizing the analysis to allow for each firm to have a different reaction coefficient to the market. This research will be published in a subsequent paper.

"Quantity Matrix"	"Firms"	""	"Firm 1"	""	"Firm 1"
"Firms"	"Strategy Mix"	"Firm"	"Cournot"	"Firm"	"Keen"
"Firm 2"	"Cournot"	1	$\frac{a - c}{3\cdot b + 2\cdot d}$	1	$\frac{a\cdot b + 2\cdot a\cdot d - b\cdot c - 2\cdot c\cdot d}{5\cdot b^2 + 10\cdot b\cdot d + 4\cdot d^2}$
"Firm 2"	"Cournot"	2	$\frac{a - c}{3\cdot b + 2\cdot d}$	2	$\frac{2\cdot a\cdot b + 2\cdot a\cdot d - 2\cdot b\cdot c - 2\cdot c\cdot d}{5\cdot b^2 + 10\cdot b\cdot d + 4\cdot d^2}$
"Firm 2"	"Keen"	1	$\frac{2\cdot a\cdot b + 2\cdot a\cdot d - 2\cdot b\cdot c - 2\cdot c\cdot d}{5\cdot b^2 + 10\cdot b\cdot d + 4\cdot d^2}$	1	$\frac{a - c}{4\cdot b + 2\cdot d}$
"Firm 2"	"Keen"	2	$\frac{a\cdot b + 2\cdot a\cdot d - b\cdot c - 2\cdot c\cdot d}{5\cdot b^2 + 10\cdot b\cdot d + 4\cdot d^2}$	2	$\frac{a - c}{4\cdot b + 2\cdot d}$

Figure 15: Output levels (symbolic) under Cournot & Keen strategy combinations

Figure 16 shows the numeric outcomes with parameter values of a=800, $b=10^{-7}$, c=100 & $d=10^{-8}$. Clearly, the Keen/Keen combination results in the lowest aggregate output, and the highest price; Cournot/Cournot gives the highest aggregate output and lowest price; while the mixed strategy results in the highest output for one firm and the lowest for the other, with an intermediate aggregate output.

"Quantity Matrix"	"Firms"	""	"Firm 1"	""	"Firm 1"
"Firms"	"Strategy Mix"	"Firm"	"Cournot"	"Firm"	"Keen"
"Firm 2"	"Cournot"	1	2.2×10^9	1	1.4×10^9
"Firm 2"	"Cournot"	2	2.2×10^9	2	2.5×10^9
"Firm 2"	"Keen"	1	2.5×10^9	1	1.7×10^9
"Firm 2"	"Keen"	2	1.4×10^9	2	1.7×10^9

Figure 16: Output levels (numeric) under Cournot & Keen strategy combinations

Figure 17 shows why firms are tempted to "defect"—or in our terms, to move from not interacting to behaving strategically at this level of analysis. The firm that reacts to its competitor and prices where marginal revenue equals marginal cost will produce a greater quantity, which is only partly offset by a lower market price—so long as its competitor does not change its strategy. It unambiguously increases its profit, while that of its competitor falls. However, the same temptation also applies to the competitor, so both are likely to switch to interacting strategically. This is the temptation that makes the Cournot/Cournot combination a Nash Equilibrium, even though it involves an unambiguously lower profit for both firms.

"Profit Matrix"	"Firms"	"Profit change"	"Firm 1"	"Profit Change"	"Firm 1"
"Firms"	"Strategy Mix"	"Firm"	"Cournot"	"Firm"	"Keen"
"Firm 2"	"Cournot"	1	$-\frac{b^2\cdot(a-c)^2}{4\cdot(2\cdot b+d)\cdot(3\cdot b+2\cdot d)^2}$	1	$-\frac{b^2\cdot(a-c)^2\cdot\left(9\cdot b^2+20\cdot b\cdot d+8\cdot d^2\right)}{4\cdot(2\cdot b+d)\cdot\left(5\cdot b^2+10\cdot b\cdot d+4\cdot d^2\right)^2}$
"Firm 2"	"Cournot"	2	$-\frac{b^2\cdot(a-c)^2}{4\cdot(2\cdot b+d)\cdot(3\cdot b+2\cdot d)^2}$	2	$\frac{b^2\cdot(a-c)^2\cdot\left(\frac{7\cdot b^2}{4}+3\cdot b\cdot d+d^2\right)}{(2\cdot b+d)\cdot\left(5\cdot b^2+10\cdot b\cdot d+4\cdot d^2\right)^2}$
"Firm 2"	"Keen"	1	$\frac{b^2\cdot(a-c)^2\cdot\left(\frac{7\cdot b^2}{4}+3\cdot b\cdot d+d^2\right)}{(2\cdot b+d)\cdot\left(5\cdot b^2+10\cdot b\cdot d+4\cdot d^2\right)^2}$	1	0
"Firm 2"	"Keen"	2	$-\frac{b^2\cdot(a-c)^2\cdot\left(9\cdot b^2+20\cdot b\cdot d+8\cdot d^2\right)}{4\cdot(2\cdot b+d)\cdot\left(5\cdot b^2+10\cdot b\cdot d+4\cdot d^2\right)^2}$	2	0

Figure 17: Change in profit (symbolic) from Keen/Keen combination

Figure 18 shows the numeric consequence, given the example parameters used.

"Profit Matrix"	"Firms"	"Profit change"	"Firm 1"	"Profit Change"	"Firm 1"
"Firms"	"Strategy Mix"	"Firm"	"Cournot"	"Firm"	"Keen"
"Firm 2"	"Cournot"	1	-5.7×10^{10}	1	-1.8×10^{11}
"Firm 2"	"Cournot"	2	-5.7×10^{10}	2	1.3×10^{11}
"Firm 2"	"Keen"	1	1.3×10^{11}	1	0
"Firm 2"	"Keen"	2	-1.8×10^{11}	2	0

Figure 18: Change in profit (numeric) from Keen/Keen combination

So far, the argument looks conclusive for the Cournot-Nash Equilibrium as the outcome of strategic interaction, and competition thus works to cause higher aggregate output and lower prices than would apply with fewer firms in the industry. Add more firms, and ultimately you converge to where price equals marginal cost—the Holy Grail of perfect competition.

The acknowledged wrinkle in this argument is that, with infinitely repeated games, the Nash equilibrium is the Keen strategy—called "collusion" or "cooperate" in the literature because, before our critique, it was believed that the only way firms could work this amount out was by acting as a cartel.[19] It's possible to "rescue" perfect competition by assuming finitely repeated games, showing that "defect" (or Keen) dominates the final play, reverse-iterating back to the second last, and by finite backwards regression arrive at "defect" as the ideal strategy for all iterations. But this is obviously weak as an analogy to actual competition, where the infinitely repeated game is closer to the reality of an indefinite future of competition—even if some competitors do exit an industry, their rivals can never know when this might happen.

[19] Of course, neoclassical economists still believe this today, and will doubtless continue believing it, given how dogma has almost always overruled logic in the development of economic theory.

Most game theorists express puzzlement with this dilemma: a strategy is dominant in a one shot, but not in a repeated game. So "collusion" (or more correctly, "non-interaction") appears dominant, and it appears that firms will tend not to compete over time.[20]

There is an additional wrinkle that possibly explains this known dilemma (and the simulation results shown in **Figure 8**):[21] while the Cournot strategy is a Nash Equilibrium, it is locally unstable, and while the Keen strategy is not a Nash Equilibrium, it is locally stable. This occurs not because of collusion, but because strategic interactions—which we might describe as a "Meta-Nash Dynamic"— lead from the Cournot equilibrium to the Keen.

One firm may benefit from a strategic change—say, Firm 1 increasing its output from that in the Keen/Keen output pair, while Firm 2 reduces it. The strategy pair would then be "increase, decrease" (or "+1/-1") and the profit outcomes "increase, decrease". In an iterative search for the profit-maximizing level, this would encourage Firm 1 to continue increasing its output, which would take it in the direction of the Cournot equilibrium. However Firm 2, having lost from that strategic combination, will change its strategy—and rather than decreasing its output further, it will increase its output. Then the strategy pair will be "increase, increase" and the profit outcomes are "decrease, decrease". As a result, both firms will change their strategy to "decrease", and head back to the Keen equilibrium.

Figure 19 illustrates this using the example parameters above.[22] The top half shows the outcome for Firm 1; the bottom half, for Firm 2; strategy settings by Firm 1 are shown by column 1, and settings by Firm 2 by row one. A strategy pair of "+1/-1" results in Firm 1 increasing profit by 333, which clearly encourages Firm 1 to continue increasing production. However, that combination causes a drop in profits of 333 for Firm 2, which will cause Firm 2 to swap strategies—say from "-1" to "+1". That will then switch the market situation to the "+1/+1" combination, where both firms suffer a fall in profits (and the fall in profits gets larger for larger output increases). Both firms are then likely to switch to reducing output. The Keen equilibrium is thus locally stable because of strategic interactions.

"Firm 1"	"-1"	"0"	"+1"
"-1"	-2.1×10^{-7}	−166.7	−333.3
"0"	166.7	0	−166.7
"+1"	333.3	166.7	-2.1×10^{-7}
"Firm 2"	"-1"	"0"	"+1"
"-1"	-2.1×10^{-7}	166.7	333.3
"0"	−166.7	0	166.7
"+1"	−333.3	−166.7	-2.1×10^{-7}

[20] We suspect that this dilemma explains the paradox that neoclassical economists, who are normally so opposed to government intervention, support "competition policy", which in effect forces firms to compete with each other.

[21] One curious feature of this simulation is that the convergence result is dependent, not on the number of firms—as neoclassical theory falsely asserts—but on the dispersal of output changes by each firm. The higher the size, relative to output, of the randomly allocated output changes, the higher the likelihood that the end result will be convergence to the Cournot equilibrium rather than the Keen. This result is reported in Keen & Standish 2006.

[22] The outcome applies so long as a>c, b<a and d<c; all these are fundamental conditions for a market to exist in the first instance. a<c, for example, would set the equilibrium market output at less than zero.

Figure 19: Profit changes for Firm 1 and Firm 2 from output changes from Keen equilibrium

The Cournot equilibrium, on the other hand, is locally unstable. **Figure 20** shows the outcomes for changes of one unit for each firm. The strategy pair "+1/-1" results in increase in profits for Firm 1 and a fall in profits for Firm 2, as it did in the Keen equilibrium. Firm 1 will then be encouraged to continue increasing production, while Firm 2 will be encouraged to switch from reducing output to increasing output. The next strategy pair is thus likely to be "+1/+1" (or some higher combination). This also causes a loss for both firms, so another switch in strategy is likely—to reducing output.

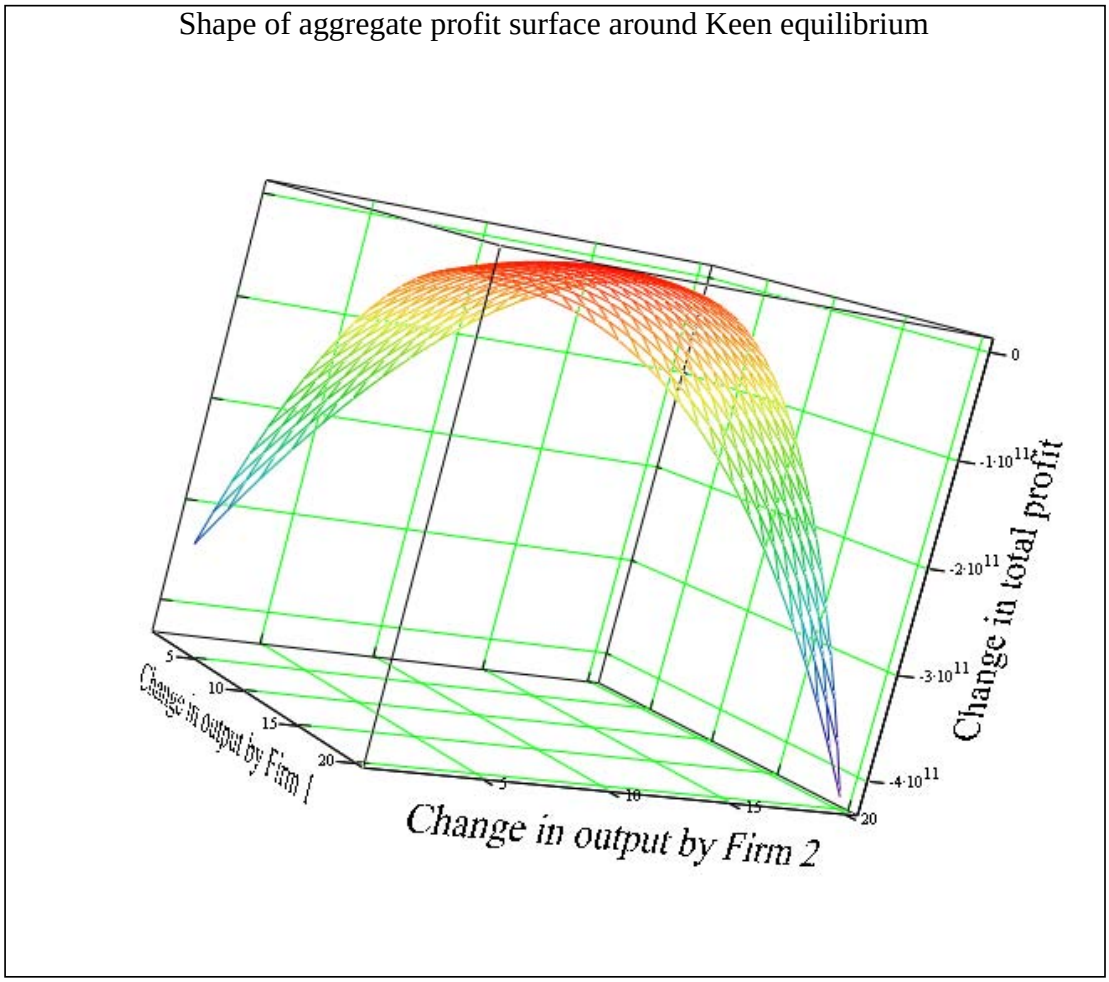
"Firm 1"	"-1"	"0"	"+1"
"-1"	218.7	-1.1×10^{-7}	–218.8
"0"	218.7	0	–218.7
"+1"	218.7	-1.1×10^{-7}	–218.8
"Firm 2"	"-1"	"0"	"+1"
"-1"	218.7	218.7	218.7
"0"	-1.1×10^{-7}	0	-1.1×10^{-7}
"+1"	–218.8	–218.7	–218.8

Figure 20: Profit changes for Firm 1 and Firm 2 from output changes from Cournot equilibrium

Unlike the Keen/Keen situation, the strategy pair "-1/-1" from the Cournot equilibrium results in an increase in profits for *both* firms—and larger reductions in output cause larger increases in profit. Further movement away from the Cournot equilibrium is rewarded, so that both firms are likely to adopt the strategy of reducing output, until they reach the Keen equilibrium—with absolutely no "collusion" taking place. The Cournot equilibrium is thus locally unstable, not because of collusion, but because of strategic interactions.

Figure 21 and **Figure 22** put the impact of strategic interactions graphically: in each case the predicted output pair (Keen/Keen and Cournot/Cournot respectively) is in the middle of the box. While firms are not behaving collusively, the only strategy pairs that have a chance to be self-sustaining are those that have a positive impact on the profit of both parties—since as explained above, any strategy that has a negative impact will necessarily mean a change in behavior by one or both firms. Therefore, the shape of the aggregate profit "hill" indicates whether any sustaining strategic interactions exist.

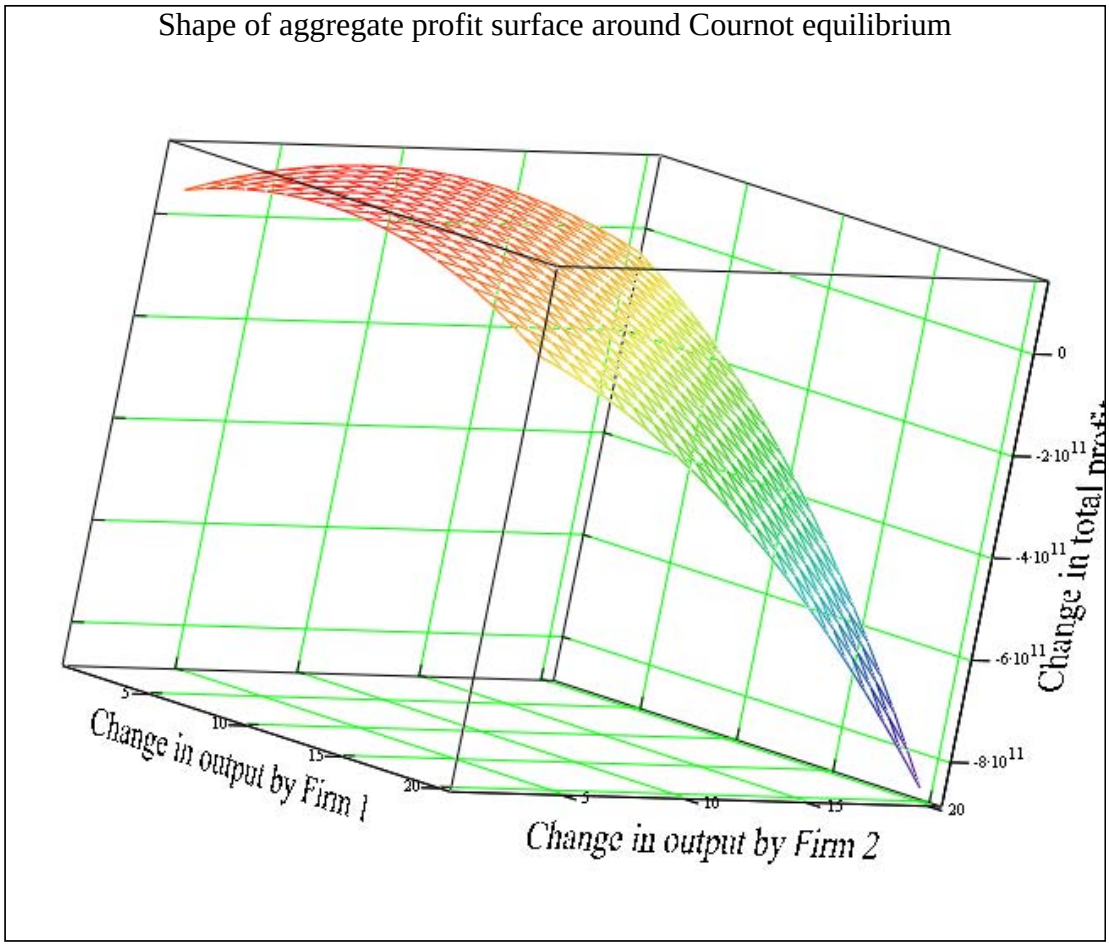
Figure 21 confirms that there are no such interactions in the vicinity of the Keen equilibrium: all strategic pairs involve a fall in aggregate profits relative to the starting point. The Keen equilibrium is thus locally stable.



M

Figure 21: Impact of strategic interactions on profit near Keen equilibrium

The Cournot equilibrium, on the other hand, is locally unstable, because aggregate profit will rise if both firms reduce output (see **Figure 22**).



N

Figure 22: : Impact of strategic interactions on profit near Cournot equilibrium

Thus, though the Cournot-Nash defense of perfect competition is not strictly false, in practice it is fragile. It appears that, if a profit-maximizing level of output per firm can be identified, then rational profit-maximizing firms will identify it, regardless of how many of them there are in an industry.[23] The Holy Grail of "perfect competition", though theoretically attainable via strategic interaction,, is a will o' the wisp.

So too, ironically, is the argument that there is a profit-maximizing level of output per firm.

The empirical reality of competition

A plethora of empirical studies have established that at least 89 per cent of output—and perhaps as much as 95 per cent—is produced under conditions of constant or falling marginal cost, and rising economies of scale.[24] Given such circumstances, there is no profit-maximizing level of output for the individual firm: so long as the sale price exceeds average costs, the firm will profit from additional sales. The key presumption of the neoclassical model—that there is a profit-maximizing level of sales—is thus not fulfilled in the real world.

[23] Subject to the one caveat mentioned in Footnote 21.

[24] See Lee 1998 for a comprehensive survey of the 20th century studies.

The most recent such survey was carried out by Alan Blinder and a team of PhD students in 1998. Blinder's results are also arguably the most authoritative, given the scale of his study, and Blinder's prestige as an economist.

Blinder et al. surveyed a representative weighted sample of US non-agricultural corporations with annual sales of more than US$10 million; a 61% response rate resulted in a study of 200 corporations whose combined output represented 7.6% of the USA's GDP. The interviews were face to face, with Blinder and a team of Economics PhD students conducting the interviews; the interviewees were top executives of the firms, with 25% being the President or CEO, and 45% a Vice President.

Blinder summarized the results in the following way:

"First, about 85 percent of all the goods and services in the U.S. nonfarm business sector are sold to "regular customers" with whom sellers have an ongoing relationship … And about 70 percent of sales are business to business rather than from businesses to consumers…

Second, and related, contractual rigidities … are extremely common … about one-quarter of output is sold under contracts that fix nominal prices for a nontrivial period of time. And it appears that discounts from contract prices are rare. Roughly another 60 percent of output is covered by Okun-style implicit contracts which slow down price adjustments.

Third, firms typically report fixed costs that are quite high relative to variable costs. And they rarely report the upward-sloping marginal cost curves that are ubiquitous in economic theory. Indeed, downward-sloping marginal cost curves are more common… If these answers are to be believed … then [a good deal of microeconomic theory] is called into question… For example, price cannot approximate marginal cost in a competitive market if fixed costs are very high." (p. 302)

The key final point about falling marginal cost deserves elaboration. Given that, as they discovered, "marginal cost is not a natural mental construct for most executives." they

translated marginal cost into 'variable costs of producing additional units," and posed the following question:

B7(a). Some companies find that their variable costs per unit are roughly constant when production rises. Others incur either higher or lower variable costs of producing additional units when they raise production.

How would you characterize the behavior of your own variable costs of producing additional units as production rises? (Blinder 1998: 102)

The survey team collated the responses into five groups, as summarized in Table 1:

Structure of Marginal Costs	Percentage of firms
Decreasing	32.6
Decreasing with discrete jumps	7.9
Constant	40
Constant with discrete jumps	7.9
Increasing	11.1

Table 1: Marginal cost structure of American corporations (Blinder et al. 1998: 102-103)

Blinder et al. pithily observed that:

> "The overwhelmingly bad news here (for economic theory) is that, apparently, only 11 percent of GDP is produced under conditions of rising marginal cost." (102)

The overall results of Blinder's survey are summarized in Table 2. Given the empirically common circumstances detailed here, the pre-requisites for being able to identify a profit-maximizing level of output do not exist for at least 89 per cent of US firms.[25] Instead, for these firms, the only profit-maximizing strategy is to sell as much as they can—and at the expense, where possible, of competitors' sales.

Summary of Selected Factual Results Price Policy	
Median number of price changes in a year	1.4
Mean lag before adjusting price months following	
Demand Increase	2.9
Demand Decrease	2.9
Cost Increase	2.8
Cost Decrease	3.3
Percent of firms which	
Report annual price reviews	45
Change prices all at once	74
Change prices in small steps	16
Have nontrivial costs of adjusting prices	43
of which related primarily to	
the frequency of price changes	69
the size of price changes	14
Sales	
Estimated percent of GDP sold under contracts	
which fix prices	28
Percent of firms which report implicit contracts	65
Percent of sales which are made to	
Consumers	21
Businesses	70
Other (principally government)	9
Regular customers	85
Percent of firms whose sales are	
Relatively sensitive to the state of the economy	43
Relatively Insensitive to the state of the economy	39
Costs	
Percent of firms which can estimate costs at least moderately well	87
Mean percentage of costs which are fixed	44
Percentage of firms for which marginal costs are	
Increasing	11
Constant	48
Decreasing	41

Table 2: Summary of Blinder et al.'s empirical findings

[25] We say at least because all previous surveys have reported a lower proportion of products that are produced under conditions of diminishing marginal productivity—typically 5 per cent of output (Eiteman & Guthrie 1952).

The only practical way that this can be done is via product differentiation, and that indeed is the obvious form that real competition actually takes. Innovation and heterogeneity are the true hallmarks of competition, yet these concepts are effectively excluded by the neoclassical model.

A model of how this actual form of competition works would be extremely useful to economic theory—and perhaps even to economic policy, if we could scientifically identify those industry structures that truly promote innovation. The continued teaching of the neoclassical model, and the continued development of a research tradition in which rising marginal cost plays a key role, are a hindrance to developing an adequate model of real world competition.

Our closing observation on this theory is perhaps the most important. A theory is more than a scholastic exercise: a good theory is also an attempt to understand reality, and, where possible, to alter it for the better. There are, therefore, few things more dangerous than an applied bad theory. Unfortunately, neoclassical competition theory is applied throughout the world, in the guise of policies intended to promote competition.

The anti-capitalist nature of neoclassical competition policy

The neoclassical vision of competition has been enshrined in competition policies adopted by governments and applied to key industries such as telecommunications, power, sanitation, and water supply. The major practical implications of accepted theory are that more firms equates to increased competition, increased competition means higher output at lower prices, and market price should ideally be equal to marginal cost.

Since the theory is flawed, these implications are at best unproven, and at worst false. There are now numerous instances around the world where competition policies have resulted in deleterious outcomes; a special issue of *Utilities Policy* in 2004 details several of these for the USA (and Australia). Loube, for example, examined the US Telecom Act of 1996, and found that “this policy has actually raised prices for residential customers” (Trebing & Miller 2004: 106).

Proponents of competition policy normally ascribe such outcomes to poor implementation of policy, poor regulatory oversight, or unanticipated circumstances. However, if the theory is flawed as we argue, then these outcomes are not accidents, but the systemic results of imposing a false theory on actual markets. Some predictable negative consequences are rising costs due to reduced economies of scale, underinvestment caused by imposed prices that lie below average cost, and reduced rates of innovation in related industries, caused by the inadequate “competitive” provision of infrastructure.

That these policies were imposed in a well-meaning attempt to improve social welfare cannot detract from the fact that, if the theory guiding these policies was false, then the policies are likely to cause more harm than good. Real world markets would function far better if competition policy, as it stands, were abolished.

Conclusion

A careful examination of the neoclassical theory of competition thus finds that it has little, if any, true content.

The Marshallian argument, which forms the backbone of neoclassical pedagogy, is strictly false in its belief that a downward-sloping market demand curve is compatible with horizontal individual firm demand curves. Once this error is corrected, the model's major conclusion, that competitive industries are better than concentrated ones, is overturned. Given identical demand and cost conditions, competitive industries will produce the same output as monopolies, and sell at the same price—and there are good grounds for expecting that monopolies would have lower costs (see Appendix One).

The Cournot analysis is mathematically correct, but subject to a problem of local instability as well as the known dilemma of repeated games. If it is interpreted as an "as if" explanation for what happens between competing firms in an industry—i.e., it proposes that firms do not actually solve the mathematics to find their Nash equilibrium output levels, but instead undertake an iterative search of the output-profit space—then this iterative search will lead to the Keen equilibrium, not the Cournot-Nash one, because the former is locally stable under strategic interactions, while the latter is not.

Given this intrinsic barrenness of the theory, its empirical irrelevance is even more important. Neoclassical economists have ignored a multitude of empirical papers that show that marginal cost does not rise, that firms do not compete on price, and so on, on the basis of Friedman's belief that asking businessmen what they do is not "a relevant test of the associated hypothesis." But if the "associated hypothesis" is in fact false, or irrelevant, then "asking businessmen what they do" is at least a good place from which to derive stylized facts that a relevant hypothesis would have to explain. It is high time that economists abandoned what superficially appears to be "high theory", and got their hands dirty with real empirical research into actual firms and actual competition.

Here the picture that emerges from even a cursory examination of the data is very different to neoclassical belief. **Table 3** shows the aggregate distribution of firm sizes in the USA in 2002: large firms make up well under 0.3 per cent of the total number of firms, but are responsible for over 60 per cent of sales.

		2002		
Industry		Total	0-499	500+
Total	Firms	5,697,759	5,680,914	16,845
	Estab.	7,200,770	6,172,809	1,027,961
	Emp.	112,400,654	56,366,292	56,034,362
	Ann. pay.($000)	3,943,179,606	1,777,049,574	2,166,130,032
	Receipts($000)	22,062,528,196	8,558,731,333	13,503,796,863

Table 3: US firm size data (US Office of Small Business Advocacy)

At the same time, small firms are not negligible: all industries are characterized by a wide distribution of firm sizes, from sole trader through to large conglomerates (see Table 4). Perhaps the real story of competition is the survival of such diversity.

	Manufacturing Firms	Estab.	Emp.	Ann. pay.($000)	Receipts($000)
Total	297,873	344,341	14,393,609	580,356,005	3,937,164,576
0 *	21,731	21,761	0	2,231,805	15,166,970
1-4	97,197	97,232	219,951	5,823,048	27,659,982
5-9	55,597	55,702	372,245	10,533,204	44,184,220
10-19	46,851	47,200	639,036	19,888,764	80,892,263
0-19	221,376	221,895	1,231,232	38,476,821	167,903,435
20-99	58,198	62,443	2,375,691	82,257,351	346,024,892
100-499	14,124	23,727	2,488,018	91,152,085	460,526,128
0-499	293,698	308,065	6,094,941	211,886,257	974,454,455
500+	4,175	36,276	8,298,668	368,469,748	2,962,710,121

Table 4: Distribution of firm sizes in manufacturing (US SBA)

In the light of both its theoretical weaknesses and its irrelevance to the empirical data, Sraffa's advice in 1930 about what to do with Marshall's theory bear repeating today, not only in relation to Marshall's theory, but even to the Cournot-Nash approach:

the theory cannot be interpreted in a way which makes it logically sell-consistent and, at the same time, reconciles it with the facts it sets out to explain. Mr. Robertson's remedy is to discard mathematics, and he suggests that my remedy is to discard the facts; perhaps I ought to have explained that, in the circumstances, I think it is Marshall's theory that should be discarded. (Sraffa 1930: 93)

The neoclassical theory of competition is a hindrance to understanding real markets and real competition, and it should be abandoned.

Appendices

Appendix One: Conditions for comparability of cost structures

Economists blithely draw diagrams like Figure 23 below to compare monopoly with perfect competition. As shown above, the basis of the comparison is false: given Marshallian assumptions, an industry with many "perfectly competitive" firms will produce the same amount as a monopoly facing identical demand and cost conditions—and both industry structures will lead to a "deadweight loss". However, in general, small competitive firms would have different cost conditions to a single firm—not only because of economies of scale spread result in lower per unit fixed costs, but also because of the impact of economies of scale on marginal costs.

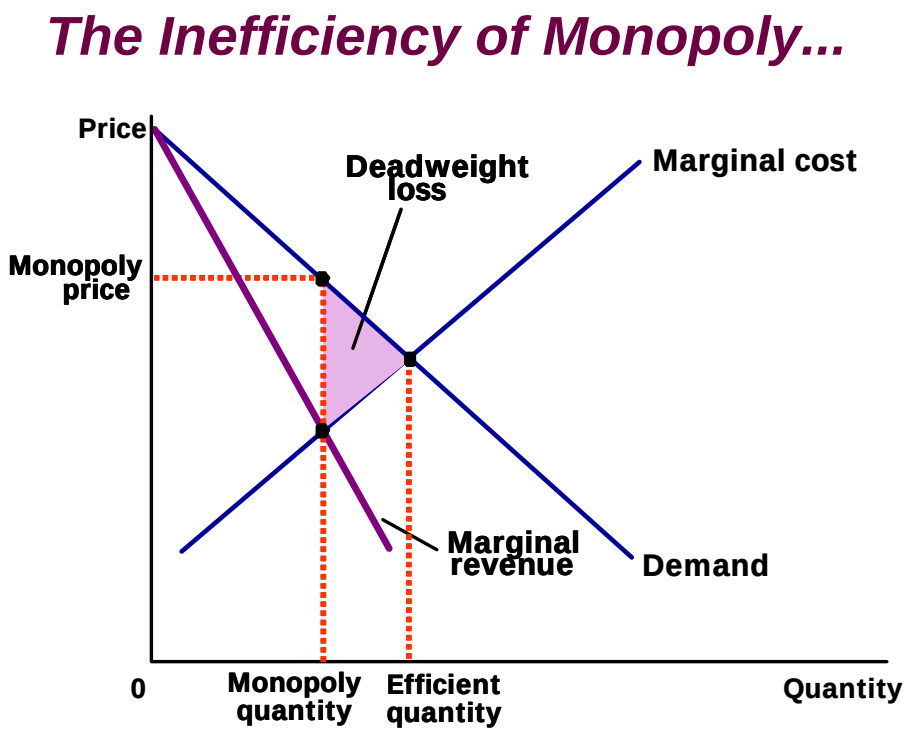


Figure 23: Mankiw's monopoly versus perfect competition comparison

Rosput (1993) gives a good illustration of this latter point in relation to gas utilities. One of the fixed costs of gas supply is the pipe; one of the variable costs is the compression needed to move the gas along the pipe. A larger diameter pipe allows a larger volume of gas to be passed with lower compression losses, so that the larger scale of output results in lower marginal costs:

Simply stated, the necessary first investment in infrastructure is the construction of the pipeline itself. Thereafter, additional units of throughput can be economically added through the use of horsepower to compress the gas up to a certain point where the losses associated with the compression make the installation of additional pipe more economical than the use of additional horsepower of compression. The loss of energy is, of course, a function of, among other things, the diameter of the pipe. Thus, at the outset, the selection of pipe diameter is a critical ingredient in determining the economics of future expansions of the installed pipe: the larger the diameter, the more efficient are the future additions of capacity and hence the lower the marginal costs of future units of output (Rosput 1993: 288).

Thus a single large supplier is likely to have lower costs—in which case, the marginal cost curve for the monopoly should be drawn *below* that for the competitive industry. Given the same demand curve and the same profit-maximizing behavior, a monopoly is thus likely to produce a higher output than a competitive industry, and at a lower cost.

The cost examples in this paper were artificially constructed to ensure that the assumption of identical costs embodied in Figure 23 were fulfilled—something that we doubt has been done by neoclassical authors in comparable papers. The cost functions were:

$$Monopoly: MC(Q) = C + D \cdot Q + E \cdot Q^2$$
$$Competitive: mc(q,n) = C + D \cdot n \cdot q + E \cdot n^2 \cdot q^2 \quad (0.20)$$

Obviously, it is very arbitrary to have the number of firms in an industry as an argument in the marginal cost function of a single firm—and also highly unlikely. Yet without that "heroic" assumption, the aggregate of marginal costs curves for a competitive industry will *necessarily* differ from the marginal cost curve for a monopoly. If a monopoly has greater access to economies of scale than smaller competitive firms, as in Rosput's example of gas transmission, then on conventional profit-maximizing grounds, a monopoly would produce a higher output for a lower price.

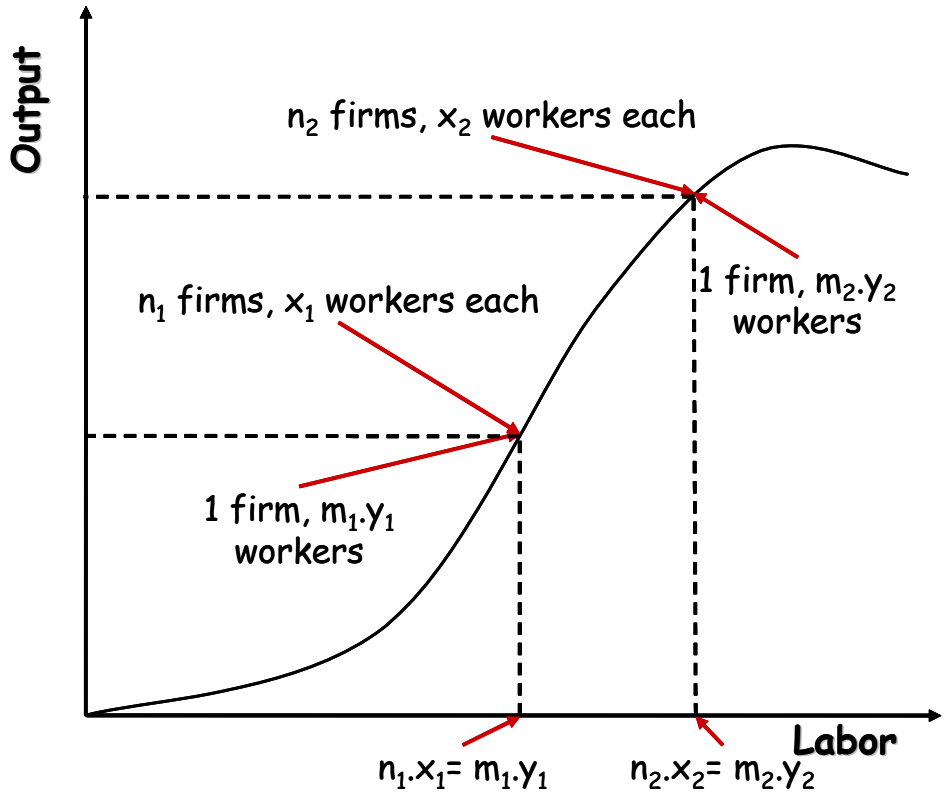
It is also easily shown that the neoclassical pedagogic assumption that the same marginal cost curve can be drawn for a competitive industry and a monopoly is true in only two circumstances: either the monopoly simply changes the ownership of plants in the industry—so that there is no technical difference between one industry structure and the other—or both industry structures face identical *constant* marginal costs.[26]

Marginal cost is the inverse of marginal product, which in turn is the derivative of total product. The condition of identical marginal costs—that is, that the marginal cost curve for a monopoly is identically equal to the sum of the marginal cost curves of an industry with many competitive firms, for all relevant levels of output—therefore requires that the total products of two different industry structures differ only by a constant. This constant can be set to zero, since output is zero with zero variable inputs.

Consider a competitive industry with n firms, each employing x workers, and a monopoly with m plants, each employing y workers, where $n>m$. Graphically this condition can be shown as in Figure 24.

[26] This argument was first published in Keen 2004a.

Figure 24: Production functions required for identical marginal cost curves



Using *f* for the production function of the competitive firms, and *g* for the production function of the monopoly, the equality of total products condition can be put in the following form:

$$n \cdot f(x) = m \cdot g(y) \qquad (0.21)$$

Substitute $y = \frac{n \cdot x}{m}$ into (1) and differentiate both sides of (0.21) by *n*:

$$f(x) = \tfrac{x}{m} \cdot g'\left(\tfrac{n \cdot x}{m}\right) \qquad (0.22)$$

This gives us a second expression for *f*. Equating these two definitions yields:

$$\frac{g\left(\frac{n \cdot x}{m}\right)}{n} = \frac{x}{m} \cdot g'\left(\frac{n \cdot x}{m}\right)$$

or

$$\frac{g'\left(\frac{n \cdot x}{m}\right)}{g\left(\frac{n \cdot x}{m}\right)} = \frac{m}{n \cdot x} \qquad (0.23)$$

The substitution of $y = \frac{n \cdot x}{m}$ yields an expression involving the differential of the log of *g*:

$$\frac{g'(y)}{g(y)} = \frac{1}{y} \qquad (0.24)$$

Integrating both sides yields:

$$\ln\left(g\left(y\right)\right) = \ln\left(y\right) + c \qquad (0.25)$$

Thus *g* is a constant returns production function:

$$g\left(y\right) = C \cdot y \qquad (0.26)$$

It follows that *f* is the *same* constant returns production function:

$$f\left(x\right) = \frac{m}{n} \cdot C \cdot \frac{n \cdot x}{m} \qquad (0.27)$$

With both *f* and *g* being identical constant returns production functions, the marginal products and hence the marginal costs of the competitive industry and monopoly are constant and identical. The general rule, therefore, is that welfare comparisons of perfect competition and monopoly are only definitive when the competitive firms and the monopoly operate under conditions of constant identical marginal cost.

References

Blinder, A.S., Canetti, E., Lebow, D., & Rudd, J., (1998). *Asking About Prices: a New Approach to Understanding Price Stickiness*, Russell Sage Foundation, New York.

Eiteman, W.J. & Guthrie, G.E., (1952). "The shape of the average cost curve", *American Economic Review* 42: 832-838.

Craig Freedman (1998). "No ends to means: George Stigler's profit motive", *Journal of Post Keynesian Economics* 20; 621-648.

Milton Friedman (1953). "The Methodology of Positive Economics", in *Essays in Positive Economics*, University of Chicago Press, Chicago: 3-43.

Haines, W.W., (1948). "Capacity production and the least cost point", *American Economic Review* 38: 617-624.

Steve Keen (1993a). "The misinterpretation of Marx's theory of value", *Journal of the History of Economic Thought*, 15 (2), Fall, 282-300.

Steve Keen (1993b). "Use-value, exchange-value, and the demise of Marx's labor theory of value", *Journal of the History of Economic Thought*, 15 (1), Spring, 107-121.

Steve Keen (2001). *Debunking Economics*, Pluto Press & Zed Books, Sydney & London.

(2004a). "Why economics must abandon its theory of the firm", in Salzano, M., & Kirman, A. (eds.), *Economics: Complex Windows*, Springer, New York, pp. 65-88.

(2004b). "Deregulator: Judgment Day for Microeconomics", *Utilities Policy*, 12: 109 –125.

(2004c). "Improbable, Incorrect or Impossible: The persuasive but flawed mathematics of microeconomics", in Fullbrook, E. (ed.), *Student's Guide to What's Wrong with Economics*, Routledge, London, pp. 209-222.

Steve Keen and Russell Standish (2005). "Irrationality in the neoclassical definition of rationality", *American Journal of Applied Sciences* (Sp.Issue): 61-68.

Steve Keen and Russell Standish (2006). "Profit Maximization, Industry Structure, and Competition: A critique of neoclassical theory", *Physica A* 370: 81-85.

Lee, F., (1998). *Post Keynesian Price Theory*, Cambridge University Press, Cambridge.

A. Mas-Colell, M.D. Whinston, J.R. Green (1995). *Microeconomics*, Oxford University Press, New York.

Piero Sraffa (1930). "Increasing Returns And The Representative Firm A Symposium", Economic Journal 40 pp.79-116.

George J. Stigler (1957), Perfect competition, historically contemplated. *Journal of Political Economy* 65: 1-17.

Harry M. Trebing and Edythe S. Miller (2004). "Introduction", *Utilities Policy* 12: 105-108.

Leon Walras (1899 [1954]). *Elements of Pure Economics*, 4th edition, George Allen & Unwin, London.

SUGGESTED CITATION:
Steve Keen and Russell Standish, "Debunking the theory of the firm – chronology", *real-world economics review*, issue no. 53, 26 June 2010, pp. 56-94, http://www.paecon.net/PAEReview/issue53/KeenStandish53.pdf

The high budgetary cost of incarceration

John Schmitt, Kris Warner, and Sarika Gupta
[Center for Economic and Policy Research, USA]

Introduction

The United States currently incarcerates a higher percentage of its population than any other country in the world. In 2008, over 2.3 million Americans were in prison or jail, and one of every 48 working-age men was behind bars. These rates are not just far above those of the rest of the world, they are also substantially higher than our own long-standing historical experience. The financial costs of our corrections policies are staggering. In 2008, federal, state, and local governments spent about $75 billion on corrections,[1] the large majority of which was spent on incarceration. Reducing the number of non-violent offenders in our prisons and jails by half would lower this bill by $16.9 billion per year, with the largest share of these savings accruing to financially squeezed state and local governments. Every indication is that these savings could be achieved without any appreciable deterioration in public safety.

This report first documents the high and rising rates of incarceration in the United States, comparing the U.S. prison and jail population to the rest of the world and to our own historical experience. The report then reviews the main causes for the rise in incarceration and analyzes the relationship between incarceration and national crime rates. The final section of the report quantifies some of the direct financial costs of incarceration and discusses the scope for budgetary savings, particularly for state and local governments.

Incarceration nation

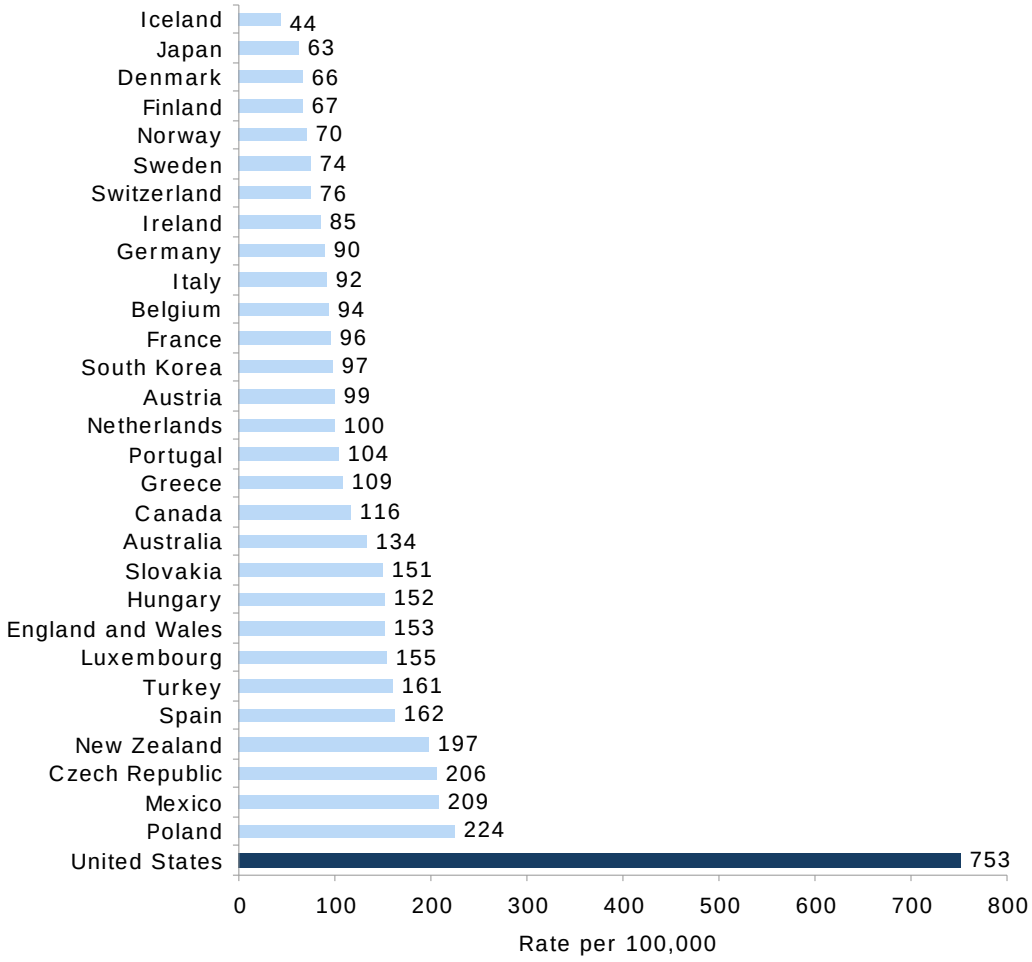
The United States has, by far, the highest incarceration rate among the rich countries that are members of the Organization for Economic Cooperation and Development (OECD). **Figure 1** shows the number of inmates per 100,000 people in the 30 OECD countries. Using the most recent data available, in the United States 753 of every 100,000 people are in prison or jail.[2] This rate is more than three times higher than the country with the next-highest incarceration rate, Poland, with a rate of 224. The U.S. rate is over seven times higher than the median rate for the OECD (102) and about 17 times higher than the rate in Iceland (44), the OECD country with the lowest incarceration rate. (**Table 1** presents the incarceration rates for the same countries for the years 1992, 1995, 1998, 2001, 2004, and the most recent year available.)

[1] Authors' projection of 2006 Bureau of Justice Statistics cost data (the most recent available), based on the increase in the correctional population from 2006 to 2008 and adjusted to 2008 dollars.

[2] Prisons generally house inmates serving sentences of at least one year, and are usually operated by the federal or state governments. Jails generally house inmates serving sentences of less than one year, and are usually operated by local governments. Alaska, Connecticut, Delaware, Hawaii, Rhode Island, and Vermont operate integrated systems that combine prisons and jails.

Figure 1

Incarceration Rate per 100,000 in OECD Countries (Most Recent Year, 2008-2009)



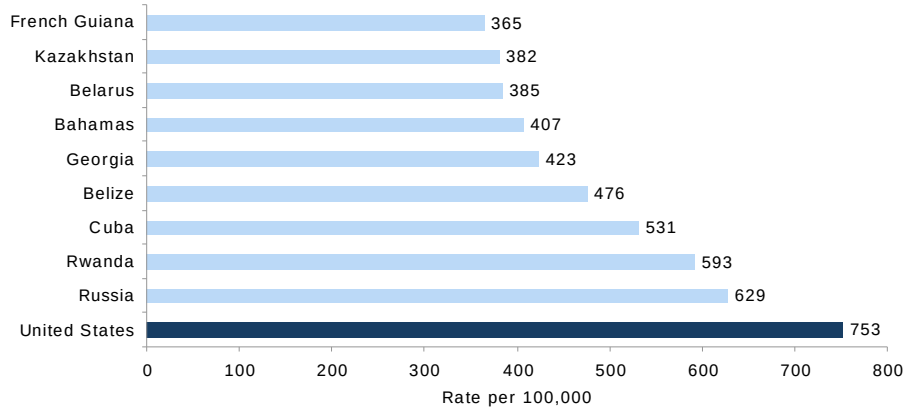
Source: Authors' analysis of ICPS data; see appendix for details.

Table 1
Incarceration Rates in OECD Countries, 1992-2008/2009

	1992	1995	1998	2001	2004	2008/2009
Australia	89	96	107	116	120	134
Austria	87	78	87	86	110	99
Belgium	71	75	82	85	88	94
Canada	123	131	126	117	108	116
Czech Republic	123	181	209	210	169	206
Denmark	66	66	64	59	70	66
England and Wales	88	99	126	127	141	153
Finland	65	59	50	59	66	67
France	84	89	86	75	92	96
Germany	71	81	96	98	98	90
Greece	61	56	68	79	82	109
Hungary	153	121	140	170	164	152
Iceland	39	44	38	39	39	44
Ireland	61	57	71	78	76	85
Italy	81	87	85	95	96	92
Japan	36	38	42	51	60	63
Luxembourg	89	114	92	80	121	155
Mexico	98	102	133	164	183	209
Netherlands	49	66	85	95	123	100
New Zealand	119	128	143	152	160	197
Norway	58	55	57	59	65	70
Poland	160	158	141	208	211	224
Portugal	93	123	144	128	125	104
Slovakia	124	147	123	138	175	151
South Korea	126	133	147	132	119	97
Spain	90	102	114	117	138	162
Sweden	63	65	60	68	81	74
Switzerland	79	80	85	71	81	76
Turkey	54	82	102	89	100	161
United States	505	600	669	685	723	753

Source: Authors' analysis of ICPS data; see appendix for details.

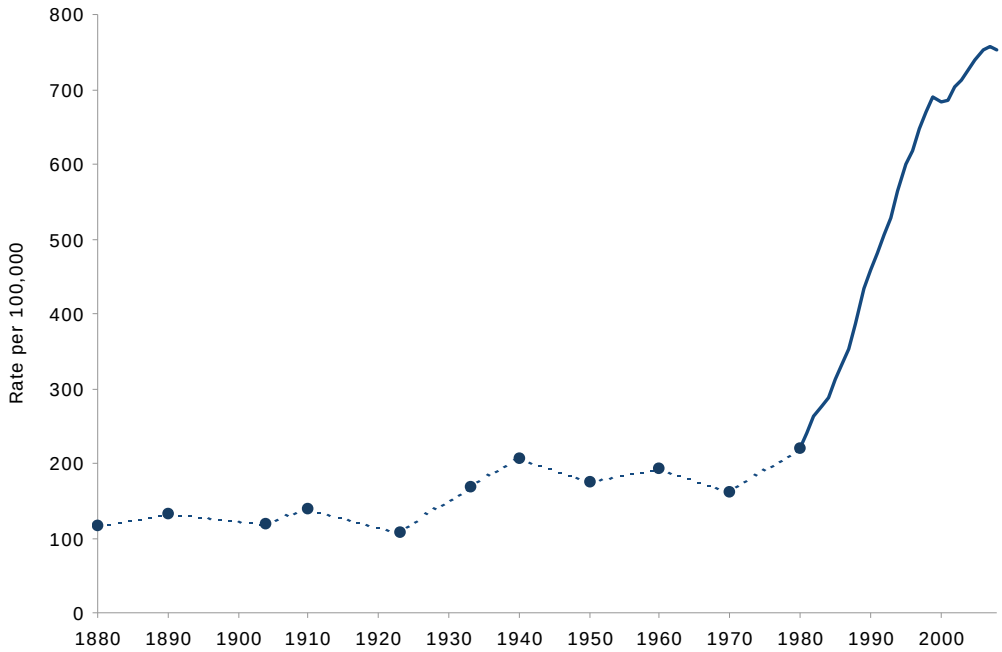
Figure 2
Top 10 Countries with Highest Incarceration Rates (Most Recent Year, 2006-2008)



Source: Authors' analysis of ICPS data, see appendix for details; excludes countries with populations less than 100,000. Data for Rwanda includes genocide suspects.

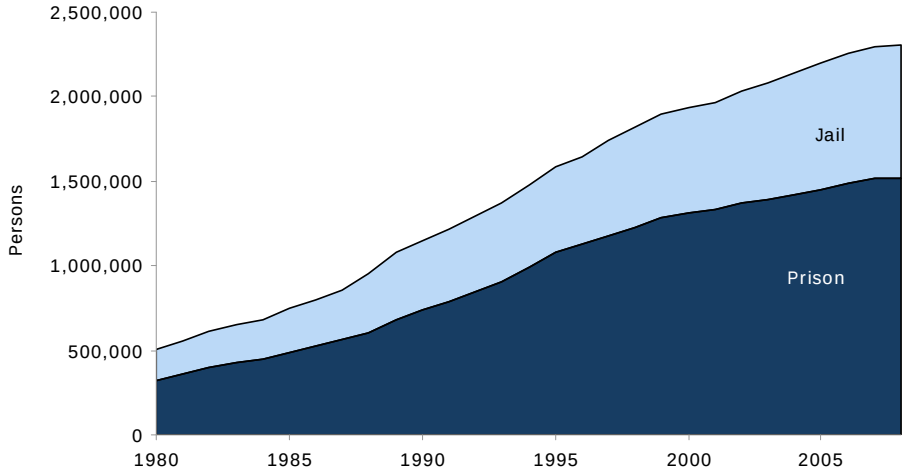
.S. incarceration rates are also high by our own historical standards. As **Figure 3** demonstrates, from 1880 to 1970 incarceration rates ranged between about 100 and 200 per 100,000.[3] From around 1980, however, the prison and jail population began to grow much more rapidly than the overall population, climbing from about 220 (per 100,000) in 1980, to 458 in 1990, to 683 in 2000, and finally to 753 by 2008. (**Figure 4** shows the total number of inmates in prisons and jails in the United States from 1980 through 2008. In 2008, there were just over 2.3 million inmates, about two-thirds in prison and about one-third in jail.)

Figure 3
U.S. Incarceration Rate, 1880-2008



Source: Bureau of Justice Statistics, Census Bureau, and Cahalan (1986). See Appendix for further details.

Figure 4
U.S. Prison and Jail Population, 1980-2008



Source: Bureau of Justice Statistics

[3] The lowest rate was 107.4 in 1923; the highest rate was 207.4 in 1940.

Table 2

Incarceration Rates for Males Age 18 to 64, 1960-2008

	1960	1980	2008
Total Prisoners	226,344	319,598	1,518,559
Prisoners, Male	217,806	302,174	1,410,260
Prisoners, Male 18-64	210,129	273,673	1,338,036
Total Jail Inmates	119,671	183,988	785,556
Jail Inmates, Male	111,866	166,305	685,790
Jail Inmates Age, Male 18-64	105,128	159,672	671,475
Total Prison and Jail, Male 18-64	315,258	433,345	2,009,512
Total US Population, Male 18-64	48,212,468	67,729,280	97,228,219
Prison and jail as percent of total US population, Males 18-64	0.65	0.64	2.07
One in every ... men age 18-64 is in prison or jail.	153	156	48

Notes: Authors' estimates based on BJS and Census data. See Appendix for details.

The standard measure of incarceration – inmates per 100,000 people in the total resident population – masks the strong concentration of men (particularly young men of color[4]) in prison and jail. Based on our analysis of Bureau of Justice Statistics data, for example, we estimate that, in 2008, 2.1 percent of working-age men, or about one in every 48 working-age men in the United States, were in prison or jail (see **Table 2**). In 1960, this figure was 1 in 153 and it changed little by 1980 when it was at 1 in 156.

Crime and punishment

Why are U.S. incarceration rates so high by international standards and why have they increased so much over the last three decades? The simplest possible explanation would be that the jump in incarceration merely reflects a commensurate rise in crime. The data, however, are clear that increases in crime since 1980 can explain only a small share of the massive rise in incarceration.

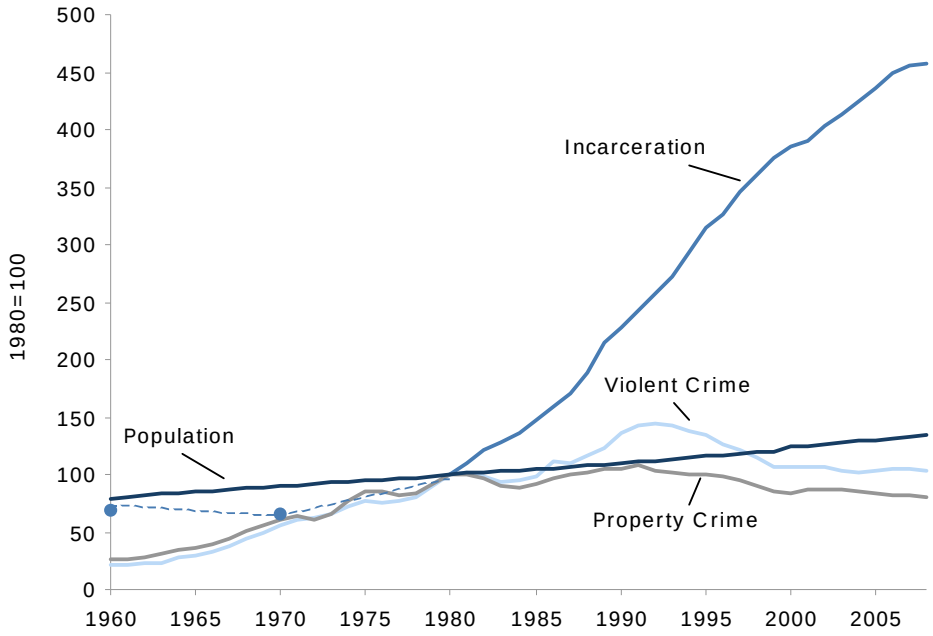
Figure 5 shows the change between 1960 and 2008 in the incarcerated population, the number of violent crimes, the number of property crimes, and the overall population. The figure sets the level of all four statistics at 100 percent in 1980 and graphs the proportional change in each measure before and after that year. The total amount of violent crime did increase after 1980, peaking in 1992 at about 44 percent above its 1980 level. Property crime

[4] For an excellent recent analysis, see Public Safety Performance Project (2008a) and Austin (2007).

also rose, but much less, peaking in 1991 at about 7 percent above its 1980 level. Over this same period that violent and property crimes were on the rise, the incarcerated population also grew, but much more rapidly – rising more than 150 percent between 1980 and 1992.

Figure 5

Change in Violent and Property Crime, and Inmate and Total Population, 1960-2008



Source: Authors' analysis of FBI and BJS data.

After 1992, both violent crime and property crime declined – returning by 2008 to close to 1980 levels in the case of violent crime and actually falling well below 1980 levels in the case of property crimes. Even as the total number of violent and property crimes fell, however, the incarcerated population continued to expand rapidly.

These data suggest that rising crime can explain only a small portion of the rise in incarceration between 1980 and the early 1990s, and none of the increase in incarceration since then. If incarceration rates, for example, had tracked violent crime rates, the incarceration rate would have peaked at 317 per 100,000 in 1992, and fallen to 227 per 100,000 by 2008 – less than one third of the actual 2008 level and about the same level as in 1980.

Stricter sentencing policies, particularly for drug-related offenses, rather than rising crime, are the main culprit behind skyrocketing incarceration rates. The last three decades have seen the implementation of new "tough on crime" policies such as three-strikes laws, truth in sentencing laws, and mandatory minimums.[5] These laws have led to a significant increase in the number people who are incarcerated for non-violent offenses. Arrests and convictions for drug offenses have increased dramatically over the last three decades,[6] with

[5] See, for example, Public Safety Performance Project (2007, 2008a), Abramsky (2007), Western (2006), Stemen, Rengifo, and Wilson (2006), and Benson (2009).

[6] See Benson (2009).

non-violent drug offenders now accounting for about one-fourth of all offenders behind bars (see **Table 3**), up from less than 10 percent in 1980.[7] Additionally, during this period, the criminal justice system has moved away from the use of probation and parole. As a result, convicted criminals today are much more likely than in the past to be sentenced to prison or jail, instead of probation, and to serve longer terms, with less chance of being released on parole.

Table 3
Inmates by Most Serious Offense, 2000s

Offense / year	Jail 2002 (Percent)	State prison 2006 (Percent)	Federal prison 2008 (Percent)	Estimated total, 2008				
				Jail	State prison	Federal prison	All inmates	
							Percent	Number
Violent Offenses	25.4	50.2	8.5	199,531	662,713	16,849	38.2	879,093
Non-violent offenses	74.5	49.8	91.5	585,239	657,432	181,567	61.8	1,424,238
Property Offenses	24.4	20.9	6.1	191,676	275,910	12,057	20.8	479,643
Drug Offenses	24.7	20.0	52.1	194,032	264,029	103,465	24.4	561,526
Public-Order Offenses	24.9	8.4	32.5	195,603	110,892	64,528	16.1	371,023
Other	0.5	0.5	0.8	3,928	6,601	1,517	0.5	12,045
Total				785,556	1,320,145	198,414		2,304,115

Notes: The estimated totals for 2008 apply the 2002 jail, the 2006 state prison, and the 2008 federal prison offense rates by type to the corresponding jail, state and federal inmate populations for 2008. Individual items may not sum to total due to rounding. Authors' analysis of Bureau of Justice Statistics, "Profile of Jail Inmates, 2002," July 2004, Table 3; and Bureau of Justice Statistics, "Prisoners in 2008," December 2009, Appendix Tables 15 and 17.

While the increase in incarceration is better explained by a shift to harsher sentencing policy than by an explosion in crime, can the case be made that higher levels of incarceration have helped to reduce crime? In a recent review of the extensive research on the relationship between incarceration and crime, Don Stemen, of the Vera Institute of Justice, concludes: "The most sophisticated analyses generally agree that increased incarceration rates have some effect on reducing crime, but the scope of that impact is limited: a 10 percent increase in incarceration is associated with a 2 to 4 percent drop in crime. Moreover, analysts are nearly unanimous in their conclusion that continued growth in incarceration will prevent considerably fewer, if any, crimes than past increases did and will cost taxpayers substantially more to achieve."[8]

Thus, the available evidence suggests that the higher rates of incarceration have made some contribution to lowering the crime rate, either by acting as a deterrent or by warehousing offenders during the ages in their lives when they are most likely to commit crimes. But, the impact of incarceration on crime rates is surprisingly small, and must be weighed against both its high monetary costs to government budgets and its high social costs to prisoners, their families, and their communities.[9]

[7] Figure for 1980 calculated based on Mauer and King (2007), who indicate that there were an estimated 41,100 drug offenders in the nation's jails and prisons in 1980.

[8] Stemen (2007). See also Austin et al. (2007), Irwi n, Schiraldi, and Ziedenberg (1999), and Public Safety Performance Project (2007).

[9] For discussion of social costs, see Austin et al. (2007).

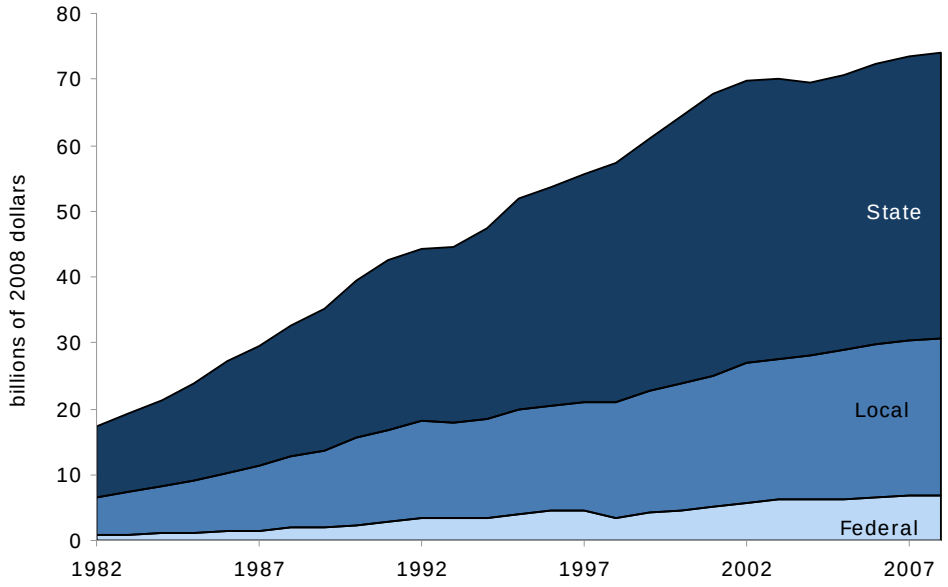
In the next and final section of this report, we examine the potential financial benefits to the federal, state, and local governments of rethinking sentencing policies for non-violent offenders.

The high cost of punishment

In 2008, federal, state, and local governments spent nearly $75 billion on corrections, with the large majority on incarceration. **Figure 6** breaks down total corrections costs across the three levels of government and illustrates that by far the largest share of the costs of corrections are borne by state and local governments. State governments house about 60 percent of inmates and account for about the same share of total correction expenditures. Local governments hold about one third of all inmates and make not quite one third of total corrections spending. The federal government, which holds less than 10 percent of the inmate population, spends just under 10 percent of total national corrections expenditures.

Figure 6

Total Inflation-adjusted Corrections Expenditures by Type of Government, 1982-2008



Source: Authors' analysis of BJS data; 2007-2008 values are estimates based on growth in respective inmate populations.

Figure 6 also demonstrates that the total cost (in inflation-adjusted dollars) of government corrections budgets has increased almost in lock-step with the incarceration rate. Given the earlier analysis that higher incarceration rates are overwhelmingly not a function of higher crime rates and that higher incarceration rates have only a small effect on crime rates, the data in Figure 6 suggest that there may be substantial scope for reducing government expenditures on corrections.

One concrete proposal for cutting expenditures on incarceration would be to reduce the number of non-violent offenders in prison and jail by half (with no change in the incarceration rates for violent offenders). **Table 4** presents the projected budgetary impact of

such a proposal, using the estimated distribution of prisoners and estimated costs for incarceration in 2008. The calculations in the table assume no change in the violent-offender population in prisons and jails, and that the reduction in non-violent-offender inmates would be largely accomplished by moving non-violent offenders to probation (for new offenders and jail inmates) or parole (current prisoners).

On the one hand, these calculations imply a large shift in corrections strategy, including moving about 700,000 current or potential inmates to probation or parole. On the other hand, despite the scale of the reduction in inmates, the overall incarceration rate would only fall from 753 per 100,000 to 521 per 100,000 and would still leave the United States with the highest incarceration rate in the OECD (more than twice the rate of second place Poland) and the fourth highest rate in the world (behind Russia at 629, Rwanda at 593, and Cuba at 531). The new implied rate of 521 per 100,000 would only return the US incarceration rate to roughly where it was in 1993 (528 per 100,000).

Table 4

Estimated Total Budgetary Savings from Reduction by Half in Non-violent Offender Incarceration

				Change in corrections spending	
	Total inmates	Change in inmates	Savings per offender	Billions of dollars	As percent of corrections budget
Federal					
Violent	16,849	0	0	0	0.0
Non-violent	181,567	-90,783	22,700	-2.1	-30.6
State					
Violent	662,713	0	0	0	0.0
Non-violent	657,432	-328,716	23,200	-7.6	-17.6
Local					
Violent	199,531	0	0	0	0.0
Non-violent	585,239	-292,620	24,700	-7.2	-30.2
Total	2,304,115	-712,119		-16.9	-22.8

Notes: Savings calculated assuming that new and existing non-violent offenders move from prison to parole or from jail to probation; that annual operating costs are $25,500 per federal prisoner, $26,000 per state prisoner and per jail inmate, $2,800 per parolee, and $1,300 per probationer, based on Public Safety Performance Project (2007, 2009) and authors' estimates. Federal, state, and local corrections budgets from Figure 6. Inmates by most serious offense type from Table 3.

The calculations in Table 4 assume that for each non-violent offender shifted from prison or jail (at an average cost of about $25,500 to $26,000 per year) to probation or parole (at average cost of $1,300 to $2,800 per year), government corrections systems would save $23,000 to $25,000 per inmate per year. Given the mix of prisoners by offense type (see Table 3), a 50 percent reduction in non-violent-offender inmates would save the federal government about $2.1 billion per year, state governments about $7.6 billion per year, and local governments about $7.2 billion per year, even after factoring in additional probation and parole costs. Across all three levels of government, these savings total $16.9 billion or about 22.8 percent of the total national spending on corrections in 2008.

Previous research by corrections officials, criminologists, and others has suggested several ways to achieve this reduction in the prison and jail population. The first is sentencing

reform. Mandatory minimum sentences, three strikes laws, and truth in sentencing laws have contributed substantially to the growing numbers of nonviolent offenders in prisons and jails.[10] Repealing these laws and reinstating greater judicial discretion would allow non-violent offenders to be sentenced to shorter terms or to serve in community corrections programs (such as parole or probation) in lieu of prison or jail.[11] Furthermore, "earned time" systems could be designed to give offenders incentives to complete education or rehabilitation programs in exchange for shorter time behind bars. [12]

Shifting non-violent offenders from prison and jail to parole and probation will be most beneficial if these community correction systems themselves are reformed. Current probation and parole policies often lack intermediate sanctions for minor violations. In many cases, the only response available for technical and minor probation or parole violations is prison or jail. A system of graduated, intermediate sanctions would give probation and parole officers and the courts authority to impose punishments short of incarceration for minor violations. [13]

Fortunately, policy makers are realizing that there are cost-effective alternatives to the status quo. The Sentencing Project, for example, notes that: "During 2009 state legislatures in at least 19 states enacted policies that hold the potential to reduce prison populations and/or promote more effective approaches to public safety."[14] At the federal level, Congress is considering legislation that would create a national commission to undertake a comprehensive examination of the criminal justice system in the United States. After an eighteen-month review period, this commission would offer recommendations "for changes in, or continuation of, oversight, policies, practices, and laws designed to prevent, deter, and reduce crime and violence, improve cost-effectiveness, and ensure the interests of justice."[15]

Conclusion

The United States has the highest incarceration rate in the world and also the highest rate in its history, with about 753 people per 100,000 in prison or jail in 2008. The number of incarcerated people in the United States has increased by more than 350 percent since 1980, while the overall population has grown by only 33 percent.

A reduction by one-half in the incarceration rate for non-violent offenders (who now make up over 60 percent of the prison and jail population) would lower the overall incarceration rate to the level reached in 1993 (which was already high by historical standards). This would also lower correctional expenditures by $16.9 billion per year, with the large majority of these savings accruing to state and local governments. These projected savings would amount to almost one-fourth of total corrections budgets. The extensive research on incarceration and crime suggests that these budgetary savings could be achieved without any appreciable deterioration in public safety.

[10] See Austin et al. (2007), Irwin, Schiraldi, and Ziedenberg (1999), and Porter (2010).

[11] See Austin et al. (2007), Porter (2010), and Public Safety Performance Project (2008a, 2009).

[12] Public Safety Performance Project (2008a, 2008b).

[13] See Austin et al. (2007), Porter (2010) and Public Safety Performance Project (2008a,2008b 2009).

[14] Porter (2010). See also Public Safety Performance Project (2008a, 2008b, 2009, 2010), Western (2008), Kleiman (2009), and Greene and Mauer (2010).

[15] Website of Senator Jim Webb (D-VA), http://webb.senate.gov/newsroom/pressreleases/2010-04-27-03.cfm.

Appendix

International incarceration rates

Figure 1

All data for incarceration rates comes from the International Centre for Prison Studies (ICPS) at King's College London. Most rates were taken from each respective country's World Prison Brief information at ICPS. The most recent rates for the following countries, however, were taken from an ICPS report entitled World Prison Population, 8th Edition: Luxembourg, Canada, Greece, Netherlands, South Korea, France, Italy, and Iceland. Also, all rates are for either 2008 or 2009 with the exception of Luxembourg whose most recent reported rate is for 2007.

Table 1

ICPS, World Prison Brief for each country, except for data in the 2008/2009 column, which is the same as the data graphed in Figure 1.

Figure 2

All data on incarceration rates were taken from each country's World Prison Brief information at ICPS, with the exception of Russia, whose rate was taken from the ICPS report entitled World Prison Population, 8th Edition. Cuba's rate is for 2006, French Guinea, 2007, and all other countries' rates, for either 2008 or 2009.

Historical incarceration rates

Cahalan (1986) provides a wealth of information on corrections, particularly incarceration, in the United States from 1850 to the early 1980s. Though data is provided from U.S. Census Bureau reports on incarcerated individuals in 1850, 1860, and 1870, those reports are not directly comparable to later reports, so we use 1880 as our starting point for the earliest statistics on incarceration rates. See **Appendix Table 1**.

Appendix Table 1

Incarcerated Population and Rates, 1880-2008

	Prison and	Rate per 100,000
1880	58,609	116.9
1890	82,239	131.5
1904	97,308	118.4
1910	128,314	138.9
1923	120,284	107.4
1933	210,418	167.5
1940	274,821	207.4
1950	264,557	174.8
1960	346,015	193.0
1970	328,020	161.4
1980	503,586	220.4
1990	1,148,702	458.2
2000	1,937,482	682.9
2008	2,304,115	753.5

Source: 1880-1970, Cahalan (1986); 1980-2008, CEPR

Working-age men in prison or jail

2008

We applied the percentage of 18-64 male prisoners under the jurisdiction of federal and state prisons to the custody figures for prisoners, both from the BJS' "Prisoners in 2008." (See below for more information on custody and jurisdictional data). The comparable report for jails, "Jail Inmates at Midyear 2008 - Statistical Tables," did not provide a similar figure; however, the report did give an estimate of males age 18 and over. We used the percentage of male prisoners age 65 and over (1.06) to derive male jail inmates age 65 and over. Data for males age 18-64 in the U.S. came from the July 1, 2008 Census Bureau estimate[16] increased by the percentage change in the overall population from July 1, 2008 to Jan 1, 2009[17] to estimate males age 18-64 as of Dec. 31, 2008 (the information provided in "Prisoners in 2008" is for Dec. 31, 2008).

1980

Prisoner and jail inmates age 18-64 are estimated from the custody totals as provided by the Bureau of Justice Statistics, using the average of the percentages for this group from 1960 and 2008. Data for males age 18-64 in the U.S. came from the April 1, 1980 Census[18] increased by the percentage change in the overall population from April 1, 1980 to Jan. 1, 1981[19] to estimate males age 18-64 as of Dec. 31, 1980.

1960

Prisoner and jail inmates age 18-64 are estimated from the (custody) totals provided in "United States Census of the Population, 1960: Inmates of Institutions," Tables 25 and 34.[20] U.S. males age 18-64 are from the April 1, 1960 Census[21] increased by half the percentage change from July 1, 1960 to July 1, 1961.[22]

Custody vs. jurisdiction counts of inmates

Custody data refer to the "physical location in which an inmate is held regardless of which entity has legal authority over an inmates" while jurisdiction data refer to "the entity having legal authority over a prisoner, regardless of where that prisoner is held."[23] Throughout this report, we have used custody rather than jurisdiction counts of prison inmates. One reason for this is that there is no jurisdiction count for inmates in jails – there is only a custody count (conducted every five to seven years in the Census of Jails, estimated in all other years through the Annual Survey of Jails, at the end of June). This means that the national jail count includes some inmates who are actually under the legal jurisdiction of a state or the federal government but are being temporarily held at a local facility. Combining custody counts from jails with jurisdiction counts from prisons would result in double counting of some inmates. Additionally, while the BJS currently provides information on the number of prisoners held in local facilities, this practice only began in its 2000 report on prisoners (for 1999 data) and so

16 http://www.census.gov/popest/national/asrh/NC-EST2008/NC-EST2008-01.xls
17 http://www.census.gov/popest/national/NA-EST2009-01.html
18 http://www.census.gov/popest/archives/1980s/stiag480.txt
19 http://www.census.gov/popest/archives/1990s/nat-total.txt
20 http://www2.census.gov/prod2/decennial/documents/41927948v2p8a-8c_ch02.pdf
21 http://www2.census.gov/prod2/decennial/documents/09768103v1p1ch6.pdf
22 http://www.census.gov/popest/archives/1990s/popclockest.txt
23 Sabol, West, and Cooper (2009), p.13.

provides only a limited time period which estimates could be derived. Jurisdiction counts for state and federal prisoners were not conducted until 1977,[24] so historical comparisons are best done using custody counts. The semi-annual and annual reports on prisoners by the Bureau of Justice Statistics contain extensive information based on jurisdictional data. Thus, when looking at only prisoners (and not the total incarcerated population) in recent years, these data can be very useful. For instance, the jurisdiction counts in these reports provide a breakdown of those prisoners who are held in custody of private (typically for-profit) facilities (see **Appendix Table 2**). In general, these prisoners are not included in custody counts.[25]

Appendix Table 2

Custody and Jurisdiction Counts of Prisoners, 1999-2008

	Custody	Jurisdiction						
					Held in Private Facilities		Held in Local Facilities	
	Total	Total	Federal	State	Subtotal	% of Total	Subtotal	% of Total
1999	1,287,172	1,363,701	135,246	1,228,455	71,208	5.2	63,635	4.7
2000	1,316,333	1,391,261	145,416	1,245,845	87,369	6.3	63,140	4.5
2001	1,330,007	1,404,032	156,993	1,247,039	91,828	6.5	70,681	5.0
2002	1,367,547	1,440,144	163,528	1,276,616	93,912	6.5	72,550	5.0
2003	1,390,279	1,468,601	173,059	1,295,542	95,707	6.5	73,440	5.0
2004	1,421,345	1,497,100	180,328	1,316,772	98,628	6.6	74,445	5.0
2005	1,448,344	1,527,929	187,618	1,340,311	107,940	7.1	73,164	4.8
2006	1,492,973	1,569,945	193,046	1,376,899	113,697	7.2	77,912	5.0
2007	1,517,867	1,598,245	199,618	1,398,627	123,942	7.8	80,621	5.0
2008	1,518,559	1,610,446	201,280	1,409,166	128,524	8.0	83,093	5.2

Source: Bureau of Justice Statistics

References

Abramsky, Sasha. 2007. *American Furies*. Boston, MA: Beacon Press. http://www.americanfuries.com

Austin, James et al. 2007. "Unlocking America." Washington, DC: The JFA Institute. http://www.jfa-associates.com/publications/srs/UnlockingAmerica.pdf

Benson, Bruce L. 2009. "Escalating the War on Drugs: Causes and Unintended Consequences." Standford Law & Policy Review 20(2). Working paper available at: http://www.coss.fsu.edu/economics/sites/coss.fsu.edu.economics/files/users/bbenson/escalating-stanford.pdf

Cahalan, Margaret Werner. 1986. "Historical Corrections Statistics in the United States, 1850-1984." Rockville, MD: U.S. Department of Justice, Bureau of Justice Statistics. http://www.ncjrs.gov/pdffiles1/pr/102529.pdf

Greene, Judith and Marc Mauer. 2010. "Downscaling Prisons: Lessons from Four States." Washington, DC: The Sentencing Project. http://www.sentencingproject.org/doc/publications/publications/inc_DownscalingPrisons2010.pdf

[24] Cahalan (1986), p. 35.

[25] Sabol, West, and Cooper (2009), p. 9.

Irwin, John, Vincent Schiraldi, and Jason Ziedenberg. 1999. "America's One Million Nonviolent Prisoners." Washington, DC: Justice Policy Institute.
http://www.justicepolicy.org/images/upload/99-03_REP_OneMillionNonviolentPrisoners_AC.pdf

Kleiman, Mark A. R. 2009. *When Brute Force Fails: How to Have Less Crime and Less Punishment*. Princeton, NJ: Princeton University Press.

James, Doris J. 2004. "Profile of Jail Inmates, 2002." Washington, D.C.: Bureau of Justice Statistics. Available at http://bjs.ojp.usdoj.gov/content/pub/pdf/pji02.pdf

Mauer, Marc and Ryan S. King. 2007. "A 25-Year Quagmire: The War on Drugs and Its Impact on American Society." Washington, DC: The Sentencing Project.

http://www.sentencingproject.org/doc/publications/dp_25yearquagmire.pdf

Minton, Todd D. and William J. Sabol. 2009. "Jail Inmates at Midyear 2008 - Statistical Tables." Washington, DC: Bureau of Justice Statistics.

National Public Radio. 2007. "Timeline: America's War on Drugs."
http://www.npr.org/templates/story/story.php?storyId=9252490

Porter, Nicole D. 2010. "The State of Sentencing 2009: Developments in Policy and Practice." Washington, DC: The Sentencing Project.
http://www.sentencingproject.org/doc/publications/s_ssr2009Update.pdf

Public Safety Performance Project. 2010. "Prison Count 2010: State Population Declines for the First Time in 38 Years." Washington, DC: Pew Center on the States, Pew Charitable Trusts.

Public Safety Performance Project. 2009. "One in 31: The Long Reach of American Corrections." Washington, DC: Pew Center on the States, Pew Charitable Trusts.
http://www.pewcenteronthestates.org/uploadedFiles/PSPP_1in31_report_FINAL_WEB_3-26-09.pdf

Public Safety Performance Project. 2008a. "One in 100: Behind Bars in American 2008." Washington, DC: Pew Center on the States, Pew Charitable Trusts.
http://www.pewcenteronthestates.org/uploadedFiles/8015PCTS_Prison08_FINAL_2-1-1_FORWEB.pdf

Public Safety Performance Project. 2008b. "Policy Framework to Strengthen Community Corrections." Washington, DC: Pew Center on the States, Pew Charitable Trusts.

Public Safety Performance Project. 2007. "Public Safety, Public Spending: Forecasing America's Prison Population, 2007-2011." Washington, DC: Pew Center on the States, Pew Charitable Trusts.
http://www.pewcenteronthestates.org/uploadedFiles/Public%20Safety%20Public%20Spending.pdf

Sabol, William J., Heather C. West, and Matthew Cooper. 2009. "Prisoners in 2008." Washington, DC: Bureau of Justice Statistics. http://bjs.ojp.usdoj.gov/content/pub/pdf/p08.pdf.

Scott-Hayward, Christine S. 2009. "The Fiscal Crisis in Corrections: Rethinking Policies and Practice." New York: Vera Institute of Justice.
http://www.pewcenteronthestates.org/uploadedFiles/Vera_state_budgets.pdf

Stemen, Don. 2007. "Reconsidering Incarceration: New Directions for Reducing Crime." New York: Vera Institute of Justice. http://www.vera.org/download?file=407/veraincarc_vFW2.pdf

Stemen, Don, Andres Rengifo, and James Wilson. 2006. "Of Fragmentation and Ferment: The Impact of State Sentencing Policies on Incarceration Rates, 1975-2002." New York: Vera Institute of Justice.
http://www.ncjrs.gov/pdffiles1/nij/grants/213003.pdf

Western, Bruce. 2008. "Reentry: Reversing Mass Imprisonment." *Boston Review*, July/August.

Western, Bruce. 2006. *Punishment and Inequality in America*. New York: Russell Sage Foundation.

SUGGESTED CITATION:
John Schmitt, Kris Warner, and Sarika Gupta, The high budgetary cost of incarceration", *real-world economics review*, issue no. 53, 26 June 2010, pp. 95-108, http://www.paecon.net/PAEReview/issue53/SchmittWarnerGupta53.pdf

Deficit hysteria redux?*

Why we should stop worrying about U.S. government deficits

Yeva Nersisyan and L. Randall Wray [Levy Economics Institute and Uni. of Missouri, K.C, USA]

Introduction

When it comes to federal budget deficits there appear to be only two respectable positions. The first is the "deficit hawk" position: deficits are never acceptable because they lead to complete crowding-out; that is, every dollar of government spending is offset by a dollar of private spending. Indeed, for the long run it is even worse, because government debt will have to be repaid in the future, which means higher taxes and less private spending. Hence, the stimulus package did not save any jobs and will actually cost us jobs later. This is a minority view among economists and policymakers, although it remains popular among some Republicans who have a political interest in denying that the Democrats and the Obama administration have done anything right.

The second view is the "deficit dove" position: deficits are probably acceptable for the short run, and perhaps even necessary to save the economy from another Great Depression. However, the benefits we receive today are partially offset by costs in the future, when we will need to tighten our belts to repay the debt. Even President Obama has argued that today's deficits will impose a heavy burden on our grandchildren, and warned that "we cannot continue to borrow against our children's future" (February 1, 2010). This is why he is already proposing budget freezes for the year after next. Other deficit doves are somewhat more tolerant of near-term budget shortfalls than the president, but they still worry about long-term pain, citing the imminent retirement of baby boomers and concomitant increase in "entitlement" spending. Thus, it is all the more necessary to get the budget "under control" as quickly as possible.

Finally, a new and influential study by Carmen Reinhart and Kenneth Rogoff (2009a) purports to show that economic growth slows dramatically—by at least one percentage point—once the gross debt–to-GDP ratio crosses the threshold of 90 percent. President Obama's proposed budget will soon cross that line, with the debt-to-GDP ratio reaching 103 percent by 2015.[1] That would drop per capita GDP growth in the United States by over half from a long-run potential of 2.5 percent per year—"the difference between a strong economy that can project global power and a stagnant, ossified society" (Boskin 2010). At that pace, living standards would rise so slowly that improvement would barely be noticed—good-bye, American dream.

In this brief, we present a third view that receives virtually no media attention. We argue that today's deficits do not burden future generations with debt that must be repaid, nor do they crowd out private spending now or in the future. The Reinhart and Rogoff findings and both of the conventional views cannot be applied to the situation of the United States, or to any other nation that operates with a sovereign currency (that is, a national currency with a floating exchange rate, and with no promise to convert at a fixed exchange rate to another currency).

* This paper originally appeared as a Levy Economics Institute of Bard College, *Public Policy Brief,* No. 111, 2010, ISBN 978-1-936192-08-3.

Our arguments are not really new—they can be found in numerous Levy Institute publications over the past two decades. Nor is the deficit hysteria new; it returns predictably on cue, like an undead monster in a horror flick, to constrain rational policy when a downturn causes the deficit to grow. In the current case, however, the stakes are higher than they have been since the 1930s. Our economy faces such strong headwinds that it requires a fiscal expansion that could result in even larger and perhaps more prolonged deficits than those now projected. Thus, it is more important than ever to explain why sustained budget deficits do not threaten our future.

Deficit and debt facts

We first present some data on federal budget deficits and debt because there is so much misinformation surrounding these measures. Budget deficits add to the outstanding stock of federal government debt. These data are often presented relative to the size of GDP, helping to scale the nominal numbers and provide some perspective. Unfortunately, this approach is often not pursued by scaremongers who talk of "tens of trillions of dollars of unfunded entitlements" when the baby boomers retire, since the figure is meaningless unless it is compared to the sum of GDP over the same period.

Figure 1 shows federal government debt since 1943 in terms of debt held by the public and gross debt. The scaremongers use the (larger) gross debt number, which is highly misleading because it includes the debt held in federal government accounts: debt the government owes itself, including securities held in civil service and military retirement funds, Social Security, Medicare, and unemployment and highway trust funds. For example, the Social Security program has run large budget surpluses since the early 1980s that are used to accumulate Treasury debt. The program will sell the bonds back to the Treasury when revenues are less than benefit payments.[2] Really, this represents internal accounting, a reminder that Social Security runs surpluses today but will run deficits in the future. The relevant debt figure is the amount of Treasuries held by the public.[3]

During World War II, the government's deficit reached 25 percent of GDP and raised the publicly held debt ratio to more than 100 percent—much higher than the 2015 forecast of 73 percent. Further, in spite of warnings by Reinhart and Rogoff (2009a), U.S. economic growth in the postwar "golden age" was robust. The debt ratio declined rather rapidly due to growth that raised the denominator of the debt ratio rather than to budget surpluses and debt retirement (although there were many small annual surpluses, as discussed below). By contrast, slower economic growth after 1973 was accompanied by budget deficits, leading to slow growth of the debt ratio, until the economic boom and budget surpluses during the Clinton years again lowered the ratio.

When federal government debt is held by the public, the government liability is exactly offset by nongovernment sector assets, and interest payments by the government generate income for the nongovernment sector. Even on the orthodox claim that today's deficits lead to debt that must be retired later, those future higher taxes that are said to be required to service and pay off tomorrow's debt represent "redistribution" from taxpayers to bondholders. This might be undesirable (perhaps bondholders are wealthier than taxpayers), but the "redistribution" takes place at the time the payment is made. While it is often claimed that deficit spending today burdens our grandchildren, in reality we leave them with government bonds that represent net financial assets and wealth. If the decision is to raise taxes and retire the bonds in, say, 2050, the extra taxes are matched by payments made directly to bondholders in 2050. (We deal with foreign holdings of government bonds below.)

Although this decision to raise taxes in an effort to retire the debt burdens taxpayers in 2050, it is not a necessary decision. If taxes are not increased later, we simply leave future generations with Treasury debt that is a net asset in their portfolios, and any payment of interest provides net income to bondholders. Obviously, it will be up to future generations to decide whether they should raise taxes by an amount equal to those interest payments or by a greater amount to attempt to retire the debt. Even if we want to, we cannot put those burdens on future generations because we cannot dictate the fiscal stance to be taken in 2050. In short, our deficits today do not necessarily commit future generations to raising taxes.

Moreover, future generations would find that their attempts to raise taxes (or slash spending) to achieve budget surpluses will fail because the budgetary outcome is mostly "endogenous" or nondiscretionary. Fiscal austerity slows the economy to the point that tax revenues fall and spending on the social safety net rises, thus preventing budget surpluses. In other words, even if future generations decide to raise taxes and burden themselves, they probably will not be able to retire the leftover Treasury debt because their actions will not ensure a budget surplus large enough to run down the debt. Recall President Clinton's promise to run surpluses for 15 years in order to retire all the outstanding debt—which failed because the fiscal drag caused a recession that restored budget deficits. Thus, our grandkids might as well enjoy the Treasuries as net wealth in their portfolios and avoid the pain of higher taxes.

That response—keeping the inherited debt—is what generations of Americans have done. There has been only one brief period in U.S. history when a generation actually imposed sufficient taxes to retire all the federal government debt: from 1835 to 1837, during Andrew Jackson's second presidential term. All other generations have adopted a much more prudent approach by growing the economy and reducing the debt ratio rather than by raising taxes or slashing spending.

The discussion so far has assumed that Treasury debt is held domestically; however, much of the current hand-wringing about deficits and debt concerns foreign ownership. Figure 2 shows foreign ownership of federal government debt as a percent of all publicly held debt. The percent held by foreigners has indeed been climbing—from less than 20 percent through the mid-1990s to nearly 50 percent today. Most growth was by "official" holders such as foreign treasuries or central banks, which now account for more than a third of all publicly held federal debt. This is supposed to represent ceding some measure of control of the U.S. government's purse strings to foreign governments. Indeed, it is frequently claimed that China is responsible for "financing" a huge portion of our federal government deficit, and that if it suddenly stopped lending, the United States might be unable to finance its budget deficits.

The U.S. current account balance largely reflects the country's trade deficit, fluctuating in the range of minus 0.5 to minus 3 percent of GDP in the 1980–99 period (Figure 2). After 1999, the balance plummeted close to minus 6 percent of GDP before turning around during the global economic downturn. However, it remains close to minus 3 percent of GDP today. Note that the rapid growth of foreign holdings of Treasuries coincided with the rapid downturn of the current account balance—a point we return to below.

Financial sector holdings of Treasuries had been on a downward trend until the current global crisis, when a run to liquidity led financial institutions to increase purchases. Financial sector holdings act like a buffer: when foreign demand is strong (weak), U.S. financial institutions reduce (increase) their share. In recent months the current account deficit has fallen dramatically and reduced the flow of dollars abroad. Of course, new Treasury issues have grown along with the rising budget deficit, and the holdings of U.S. financial

institution initially increased in the run to liquidity during the crisis. Foreign official holdings have also continued to climb, perhaps because the U.S. dollar is still seen as a refuge and nations want to accumulate dollar reserves to protect their currencies. This is the other side of the liquidity-crisis coin. If there is fear of a run to liquidity, exchange rates of countries thought to be riskier than the United States face depreciation. It is rational for any country trying to peg its currency to the dollar to increase its official holdings in response to a global financial crisis.

Figure 3 shows the foreign holdings of U.S. Treasuries. While most public discussion has focused on China, Japanese holdings have been comparable, and even surpassed those of China in December 2009. As discussed above, there is a link between current account deficits and foreign accumulation of U.S. Treasuries. From the point of view of holders, current account surpluses allow them to accumulate dollar-denominated assets. In the first instance, a trade surplus leads to dollar reserve credits (cash plus credits to reserve accounts at the Fed). Since these credits earn a low interest rate (indeed, until recently they earned no interest), they are traded for U.S. Treasuries and other asset earnings. Thus, it is not surprising to find a link among U.S. trade deficits, foreign trade surpluses, and foreign accumulation of U.S. Treasuries.

While this is usually presented as foreign “lending” to “finance” the U.S. budget deficit, one could just as well see the U.S. current account deficit as the source of foreign current account surpluses in the form of U.S. Treasuries. Indeed, as discussed above, a trade surplus against the U.S. allows a nation to accumulate dollar reserves at the Fed. These can then be traded for U.S. Treasuries, an operation that is equivalent to transferring funds from a “checking account” (reserves) at the Fed to a “savings account” (Treasuries) at the Fed. And when interest is “paid” on Treasuries, this is just a credit of dollars to that “savings account”. In a sense, it is the willingness of the United States to simultaneously run trade and government budget deficits that provides other countries the wherewithal to “finance” the accumulation of Treasuries. It is highly misleading to view this as “lending” to the U.S. government—as if the dollars spent by the federal government originate overseas.

Obviously, there must be willingness on all sides for this to occur, and most public discussion ignores the fact that China’s eagerness to run a trade surplus with the United States is linked to its hunger for dollar assets. At the same time, the U.S. budget deficit helps to generate domestic income that allows private consumption—by fueling imports and providing income for foreigners to accumulate dollar savings, even while generating Treasuries that are accumulated by foreigners. In other words, these are not independent decisions. It makes no sense to talk of China “lending” dollars to the United States without also taking into account China’s desire for net exports. Indeed, the following matters are all linked (possibly in complex ways): the willingness of China to produce for export and to accumulate dollar-denominated assets, the shortfall of Chinese domestic demand that allows the country to run a trade surplus, the willingness of Americans to buy foreign products, the high level of U.S. aggregate demand that results in a trade deficit, and the factors behind a U.S. government budget deficit. And, of course, it is even more complicated than this, since other nations, as well as global demand, are also involved.

While there are claims that China might suddenly decide against further accumulations of U.S. Treasuries, it is likely that many other relationships would also need to change to enable that to happen. For example, China might choose to accumulate euros, but there is no equivalent to the U.S. Treasury in Euroland. China could accumulate the euro-denominated debt of individual governments—say, Greece—but this debt has different risk ratings and insufficient volume to satisfy China. Further, Euroland taken as a whole (especially Germany, its strongest member) constrains domestic demand in order to run trade

surpluses. In other words, Euroland does not represent a huge net demand for global exports. If the United States is a primary market for China's excess output but euro assets are preferred over dollar assets, then exchange rate adjustment between the dollar and the euro could destroy China's market. Hence, it is not likely that China would continue to export to the US but would accumulate euro assets rather than dollars.

We are not arguing that the current situation will go on forever, although we do believe it will persist much longer than presumed by most commentators. We are instead pointing out that changes are complex, and that there are strong incentives against the sort of simple, abrupt, and dramatic shifts posited as likely scenarios. We expect that the complex linkages between balance sheets and actions will ensure that transitions are moderate and slow.

The final topic to be addressed in this section concerns interest rates (yields) and the price of Treasuries. Figure 4 shows (daily) yields on Treasuries of different maturities. Fearing inflation and possibly attempting to quell the real estate bubble, the Fed began to raise interest rates in 2004. Note that lending in the Fed funds market or purchasing federal government debt of shortest maturity represents nearly perfect substitutes from the point of view of banks. Hence, raising the Fed funds target leads to an immediate and nearly equivalent increase in yield for the shortest maturity. Determining other rates is more complex, as shorter maturities track the Fed funds rate more closely, while longer maturities may not move in tandem with the target rate.

What we see in Figure 4 is a shocking convergence of yields across the maturity structure when the Fed pushed overnight interest rates toward 5 percent. If the Fed really was trying to increase mortgage and other longer-term rates, the market completely defeated its effort. When the Fed tightened, the 10-year bond rate, which usually tracks 30-year fixed mortgage rates fairly well, moved only 100 basis points. While many have blamed the Fed for the real estate bubble because it supposedly kept rates too low, the figure shows that the Fed raised short-term interest rates sharply but their action did not result in higher long-term rates. When the crisis hit, the Fed quickly lowered short-term interest rates, but long-term rates refused to decline by much. This reflects the "run to liquidity" that is a feature of all financial crises. (Long-term rates finally did begin to decline at the beginning of 2009, before turning upward once again.) Even as short-term rates approached the lower bound of zero, long-term rates remained high, resulting in wide spreads.

Figure 5 shows a longer time frame, with the spread between one-month T-bills and long-maturity bonds widening to approximately 400 basis points in 1993, mid-2003, and today. Also shown in this figure is government debt in terms of both gross and public Treasury debt as a percent of GDP. There is some tendency for spreads to widen when the outstanding debt stock is growing relative to GDP. To be sure, the correlation is not tight, but it is suggestive. It certainly appears that the decision in early 2006 to reissue 30-year debt might have been a mistake, since recent issues of longer-term debt have been met by stubbornly high yields. (In a later section we discuss how the Treasury can easily avoid pressuring longer term rates by always sticking to short-term maturities.)

How sustainable are budget deficits?

President Obama has warned that projected budget deficits could leave our grandchildren with a "mountain of debt" unless we bring the budget under control. Gregory Mankiw (2010), who reflects the deficit-dove position, argues that, while "a budget deficit, even a large one, is called for" at times, the trouble is that Obama's budget "fails to return the

federal government to manageable budget deficits, even as the wars wind down and the economy recovers from the recession." He goes on to argue that "the president seems to understand that the fiscal plan presented in his budget is not sustainable," which is why Obama has created a commission to come up with a way to "stabilize the debt-to-GDP ratio at an acceptable level once the economy recovers."

Catherine Rampell (2010) goes further, arguing that current incentives lead myopic politicians to "see fiscal profligacy as a prerequisite for re-election," and that only by properly aligning "the interests of the country with those of the politicians who are guiding it" can we put some spine into budgeting. A commission will not be enough: Congress has imposed various rules on itself, such as pay-go, Gramm-Rudman, and the Byrd Rule, and none were able to prevent unsustainable deficits. Hence, Rampell proposes to "delegate fiscal policy—that is, taxing and spending—to a separate technocratic body, which can rely on legal authority rather than popularity." The Fed might serve as a model—a depoliticized, independent body not subject to democratic pressures. Or, perhaps even better, says Rampell, directly change the incentives of politicians so that they would be "barred from running for re-election, and even thrown in jail" if they "overspend."

Obviously, all such critiques are based on the supposition that projected future deficits—if not those we already have today—are too large to be sustained. Various indicators have been proposed: the debt-to-GDP ratio (Reinhart and Rogoff suggest a limit below 90 percent, while the Maastricht criteria impose 60 percent) or ensuring that the government debt service does not grow faster than GDP.

We can identify three financial positions related to borrowing by households or firms—what Hyman P. Minsky called hedge, speculative, and Ponzi. A hedge position can cover interest and principal payments out of income flows. A speculative position can cover interest payments only, so the principal is rolled over. A Ponzi position must "capitalize" interest, meaning that the unit must borrow just to pay interest. Some want to extend such a classification to government. If we define government "income" as tax revenue, then a speculative position would be one in which tax revenue covers all current spending, including interest payments, but debt cannot be retired—the definition of a balanced budget. However, new debt could be issued each year, so long as additional interest payments plus additional government spending increase only as fast as government tax revenue "income." In this way, government could use its capital account to issue debt and "pay for" investment-type spending.

This is a common "deficit-dove" proposal, whereby government acts like a firm by keeping a separate capital account. Here, the "sustainability" condition would depend on the relation between the interest rate paid and the growth rate of tax revenue and other spending but would allow the government debt to grow at the growth rate of GDP. Conceivably, the debt-to-GDP ratio could even rise for some time if taxes grew faster than GDP (although taxes would eventually reach 100 percent of GDP--not a sustainable trend). For an infinitely-lived economic unit, a speculative position would appear to be safe, although rising interest rates or a fall in tax revenues and increased spending on the social safety net in a recession could turn a speculative position into a Ponzi position by producing large deficits. As Mankiw warns, current budget projections show a rising debt-to-GDP ratio and perhaps unrealistically optimistic forecasts of economic growth. Further, should the economy begin to recover, it is almost certain that the Fed would begin to raise interest rates—increasing federal spending on interest. Hence, it looks like Ponzi finance is a likely outcome of the current fiscal stance. In that case, government would "borrow" in order to finance at least some of its interest payments, and would be unable to repay its debt.

In the next section we examine whether it is appropriate to apply such classifications to a sovereign government. In short, is there anything wrong with "Ponzi finance" by the U.S. government? We will conclude that these classifications of financial positions do not apply to the sovereign issuer of the currency. In short, we argue that government is not like a household or firm.

Is a government like a household or a firm?

Discussion of government budget deficits often begins with an analogy to household budgets: "no household can continually spend more than its income, and neither can the federal government." On the surface that might appear sensible; dig deeper, and it makes no sense at all. A sovereign government bears no obvious resemblance to a household or a firm.

First of all, the U.S. federal government is 221 years old, if we date its birth to the adoption of the Constitution. Arguably, that is about as good a date as we can find, since the Constitution established a common market in the United States, forbade states from interfering with interstate trade (for example, through taxation), gave the federal government the power to levy and collect taxes, and reserved the power to create money, regulate its value, and fix standards of weight and measurement—from whence our money of account, the dollar, comes from—for the federal government.

No head of household has such a long lifespan. This might appear irrelevant, but it is not. When you die, your debts and assets need to be assumed and resolved. Firms can be long lived, but when they go out of business or are acquired, their debts are also assumed or resolved. However, there is no "day of reckoning" or final piper-paying date for the sovereign government. True, not all governments last forever, and sometimes new governments will choose to honor the debts of "deceased" governments. But honoring debts is a choice, since a sovereign government is, well, sovereign.

Note also that in spite of all the analogies drawn between governments and households, and in concert with the statement that debts cannot be allowed to grow forever, corporations that are going concerns can and do allow their outstanding debt to grow year-over-year, with no final retirement of debt unless the firm goes out of business. In other words, long-lived firms do indeed spend more than their incomes on a continuous basis. The key, of course, is that they attempt to balance their current account and keep a separate capital account. So long as firms can service their debt, the debt can always be rolled over rather than retired. This is why some deficit doves advocate capital accounts for government. We will make a stronger argument: even the infinitely-lived corporation is financially constrained, while the sovereign, currency-issuing government is not subject to the same constraints.

Second-and far more important--households and firms do not have the power to levy taxes, issue currency, or demand that taxes be paid in the currency they issue. Rather, households and firms are users of the currency issued by a sovereign government. Both households and firms do issue liabilities, and some of these liabilities can to varying degrees fulfill some functions of "money." For example, banks issue demand deposits, which are the banks' liability that can be used by households or firms a medium of exchange, a means of debt retirement, or a store of value. However, all of these private "money things" (bank deposits or other private IOUs) are denominated in dollars, and only the sovereign government of the United States has the constitutionally provided right to fix standards of weight and measurement—that is, to name the dollar money of account.

There is no need to interpret this too narrowly. It is clear that U.S. residents can voluntarily choose to use foreign currencies or even idiosyncratic measures of worth in transactions (local currency units such as the Berkshares in the Northeast). But when all is said and done, the ability of the U.S. government to impose dollar taxes and other obligations (e.g., fees and fines), and to require that those taxes and obligations be paid in dollars, gives priority to the use of dollars (and to the denomination of most transactions and obligations in dollars) within its sovereign territories that no other currency enjoys.

Third, with one brief exception the federal government has been in debt every year since 1776. For the first and only time in U.S. history, the public debt was retired in January 1835 and a budget surplus maintained for the next two years, in order to accumulate what President Jackson's Treasury secretary, Levi Woodbury, called "a fund to meet future deficits." In 1837, the economy collapsed into a deep depression and drove the budget into deficit, and the federal government has been in debt ever since.

There have been seven periods of substantial budget surpluses and debt reductions since 1776. The national debt fell by 29 percent from 1817 to 1821, and was eliminated in 1835 (under President Jackson); it fell by 59 percent from 1852 to 1857, by 27 percent from 1867 to 1873, by more than 50 percent from 1880 to 1893, and by about a third from 1920 to 1930. Of course, the last time we ran a budget surplus was during President Clinton's second term.

Has any household been able to run budget deficits for approximately 190 out of the past 230-odd years and accumulate debt virtually nonstop since 1837? As discussed above, there are firms that grow their debt year-over-year, so it is conceivable that one might have a record of "profligate" spending to match that of the federal government. Still, the claim might be that firms go into debt to increase productive capacity and thus profitability, while government spending is largely "consumption." This seems to be why the analogy is usually made between government and household budgets. But even if it is true that households do not run persistent budget deficits for years on end, it is empirically true that the U.S. government does.

Fourth, the United States has also experienced six periods of depression that began in 1819, 1837, 1857, 1873, 1893, and 1929. Comparing these dates with the periods of budget surpluses, one finds that every significant reduction of the outstanding debt, with the exception of the Clinton surpluses, has been followed by a depression, and that every depression has been preceded by significant debt reduction. The Clinton surpluses were followed by the Bush recession that was ended by a speculative, private-debt fueled euphoria, and was followed in turn by our current economic collapse. The jury is still out on whether we might yet suffer another Great Depression. While we cannot rule out coincidences, seven periods of surplus followed by six and a half depressions (with some possibility for making it a perfect seven) should raise eyebrows. And, as we show below, our less serious downturns in the postwar period have almost always been preceded by reductions of federal budget deficits. This brings us to an obvious point: the federal government is big, and any movement in its budget position has a big impact on the economy, which is the subject of the next section. As we will discuss, the government's budget plays an important balancing role in the economy—filling demand gaps that allow the nongovernment sectors to achieve the surplus that they normally desire. For this reason, trying to operate the federal government's budget as if it were a household that normally wants to save has a disastrous impact on the economy.

Finally, the most important point is that the U.S. federal government is the sole issuer of the dollar, which is always accepted in payment. Government actually spends by crediting

bank deposits (and simultaneously crediting the reserves of banks)—as such, it can never run out of dollars. These topics are explored in detail below. But the point is that no household (or firm) is able to spend by crediting bank deposits and reserves or by issuing currency. Households and firms can spend by going into debt, but the debt must be serviced with another debt—usually a bank debt. A sovereign government makes payments (including interest payments on its debt) by issuing its own nonconvertible IOU. This is why we ultimately conclude that the notion of "Ponzi finance" does not apply to government because, unlike private debtors, it can always service its debt by crediting accounts. This is a key to understanding why perpetual budget deficits are "sustainable" in the conventional sense of that term.

We realize that distinguishing between a sovereign government and a household does not put to rest all deficit fears. But since this analogy is invoked so often, it is useful to lay out some of the important differences. When someone claims that government budget deficits are unsustainable and that the government must eventually pay back all of that debt, ask him why the U.S. government has managed to avoid retiring debt since 1837—is 173 years long enough to establish a "sustainable" pattern? Ask whether we might be able to go another 173 years without the government going "bankrupt," even though it will run deficits in most years. We do admit that historical analysis is not sufficient, since the United States today is not like the country it was in 1837. However, for reasons we will discuss, the fiscal situation faced by the U.S. government is far more favorable now than it was between 1873 and the early 1970s, when the U.S. was usually operating with a convertible currency.

In the next two sections we present an alternative view of budget deficits and then compare a sovereign country like the United States with a country operating with a nonsovereign currency, like Greece on the euro or Argentina under the currency board. We conclude that the situation faced by the United States today is indeed different from that when the nation was founded, and in all succeeding years up to the collapse of Bretton Woods in 1973. The changes that year actually removed any question of the sustainability of federal budget deficits. Once the United States eliminated the last vestiges of the gold standard, government finance entered a completely new paradigm.

How a sovereign government really spends

Governments worldwide have inflicted so many self-imposed constraints on public spending that it has become really hard to see the truth behind public spending. Naturally, we tend to think that a balanced budget for a household or government is a good thing, but we fail to make the distinction between a currency issuer and a currency user—between a sovereign and a nonsovereign country (in the monetary sense). A country that pegs its currency exchange rate to another currency (or metal) doesn't have monetary sovereignty as we define it, since its domestic policy space is constrained by the necessity to maintain the peg. What we define as a sovereign currency is one that is not pegged, meaning the government does not promise to exchange its currency at a fixed exchange rate. In that sense, a sovereign currency is not convertible. The United States, like many other developed and developing countries, has been operating on a sovereign monetary system ever since it went off the gold peg in 1973.

The key insight is that if a government issues a currency that is not backed by any metal or pegged to another currency, then it cannot be constrained in its ability to "finance" spending because it doesn't need tax and bond revenues in order to spend. Indeed, we argue that modern sovereign governments spend by crediting bank accounts—they do not really spend tax revenue nor do they borrow by selling bonds. Rather, taxes result in debits of bank

accounts. Budget deficits lead to net credits to bank accounts and budget surpluses lead to net debits. At the macroeconomic level, government spending increases private disposable income while taxes reduce spending. A deficit occurs when the government adds more to private disposable income than it taxes away. A government surplus (deficit) has to equal the nongovernment sector's deficit (surplus).

Government normally sells Treasuries more or less equal in volume to its budget deficit, for reasons explained below. As already stated, budget deficits generate nongovernment surpluses or saving because government spending in excess of taxes creates nongovernment income. When the Treasury sells bonds, some of the income created by its deficits is devoted to saving in the form of government bonds. Otherwise, this saving can be held in the form of noninterest-earning cash or bank deposits. When the value of Treasury checks to the private sector is greater than the value of private sector checks that pay taxes, the private sector receives net income and accumulates wealth—in the form of Treasuries.

Private banks, in turn, accumulate Treasuries plus reserves at the Fed in an amount equal to the government's deficit (less any cash withdrawn). We can think of reserves at the Fed as the equivalent of bank "checking deposits"—used by banks for clearing (with each other and with the Fed). Treasuries can be thought of as bank "saving deposits" held at the Fed, earning interest. When a bank buys a Treasury bond, the Fed debits the bank's "checking account" (reserves) and credits the bank's "saving account" (Treasuries). The Fed credits interest to bank "saving accounts" (Treasuries)—and now also credits bank "checking accounts" with interest because the Fed started paying (low) interest on reserves last year.

When the Treasury spends by sending a check to the private sector, the check is deposited in a bank. (Increasingly, deposits are made directly by wire transfer.) The Fed then credits the bank's reserve account and debits the Treasury's account at the Fed. The opposite procedure happens when the public pays taxes: the Treasury's account at the Fed is credited and the bank's reserve account is debited, along with the taxpayer's deposit.

In case the public decides it doesn't want bank deposits and would rather have cash, households and firms withdraw currency from their bank accounts, and bank reserves decrease by an equal amount. The same happens when the public prefers to keep its wealth in the form of government bonds. The sale of Treasuries to the public results in a debit in the banks' reserve account as bond buyers write checks against their bank accounts, and the Fed debits the reserve accounts of banks and credits the Treasury's account at the Fed.

Every time the Treasury spends, bank reserves are credited, as long as the nonbank sector does not withdraw cash from its accounts. If banks already have the quantity of desired reserves (which would be the normal case), Treasury spending creates excess reserves in the system. Banks offer excess reserves in the overnight Fed funds market. Of course, all this can do is to shift the excess reserves from one bank to another, since reserves will not leave the banking system except through cash withdrawals. Finding no takers for the reserves, this will place downward pressure on the Fed funds rate, unless the Fed intervenes.

In order to provide a substitute for the excess reserves and hit its target rate, the Fed sells Treasuries to the private sector, thereby transforming the wealth held in the form of bank deposits and reserves into Treasury securities. Bank reserves are debited by the amount of Treasuries sold (whether banks or their customers buy them). In essence, this is a substitution of lower-earning excess reserves ("checking accounts" at the Fed) for higher-earning Treasuries ("saving accounts" at the Fed), and it is done to accommodate the demand for Treasuries as indicated by a falling overnight Fed funds rate (because banks do not want to hold the existing quantity of reserves). In other words, sales of Treasuries should

be thought of as a monetary policy operation that accommodates portfolio preferences of banks and their customers—much as banks will accommodate a desire by the public to switch funds from their checking accounts to their saving accounts.

To recap, a government deficit generates a net injection of disposable income into the private sector that increases saving and wealth, which can be held either in the form of government liabilities (cash or Treasuries) or noninterest-earning bank liabilities (bank deposits). If the nonbank public prefers bank deposits, then banks will hold an equivalent quantity of reserves, cash, and Treasuries (government IOUs), distributed according to bank preferences.

A government budget surplus has exactly the opposite effect on private sector income and wealth: it's a net leakage of disposable income from the nongovernment sector that reduces net saving and wealth by the same amount. When the government takes more from the public in taxes than it gives in spending, it initially results in a net debit of bank reserves and a reduction in outstanding cash balances. If banks had previously held the desired amount of reserves and cash (which would be the normal case, since otherwise there would have been pressure on the overnight rates), a budget surplus would result in a shortfall of reserves and vault cash. Banks could go to the Fed funds market to obtain reserves, but in this scenario there is a system shortage that cannot be met by interbank lending. As a result, a shortage of cash and reserve balances forces the private sector to sell Treasuries to the Fed in order to obtain the desired reserves. The Fed then adds reserves to the bank "checking deposits" and debits bank "saving deposits", simultaneously reducing the Treasury's deposit at the Fed and retiring the Treasuries. This retirement of government debt takes place pari pasu with government surpluses.

The three balances and the impact of government surpluses

The most recent period of federal government surpluses was the (highly extolled) Clinton surpluses from the first quarter of 1998 through the second quarter of 2001. For reasons that should now be clear, these surpluses destroyed nongovernment sector income and wealth, forcing households to borrow in order to maintain living standards. Since the United States ran current account deficits over that period, it was necessary for the (domestic) nongovernment sector to run even larger deficits to match the government's surplus, plus the foreign sector current account deficit.[4]

Household borrowing accelerated in the decade following the surpluses of 1998, increasing from 67 percent to 97 percent of GDP by 2007. By contrast, household debt increased from just 40 percent to 65 percent of GDP over the entire 1960–97 period. The story wouldn't be complete without predatory lenders, who were eager to extend credit to everyone, regardless of the ability to repay; and deregulation, which freed the lenders' hands (topics beyond the scope of this brief).

Based on the work of Distinguished Scholar Wynne Godley, it is useful to divide the macroeconomy according to its three main sectors: domestic government, domestic nongovernment (or private), and the foreign sector. According to his aggregate accounting identity, the deficits and surpluses across these three sectors must sum to zero; that is, one sector can run a deficit so long as at least one other sector runs a surplus. Figure 6 shows the three main balances of the United States. When there is a government sector surplus as well as a current account deficit (the "twin leakages" during the Clinton boom), the private sector is left with two possibilities to counteract the destruction of income: it can stop importing (leading to a balanced current account) or it can spend more (running a private sector deficit). For

most households, borrowing substituted for the income squeezed by tight fiscal policies. This is why the federal budget surpluses under Clinton did not (immediately) lead to an economic downturn, since private sector deficits maintained aggregate demand but increased indebtedness.

As evidenced by the current crisis, private sector borrowing on the scale seen after 1997 is not sustainable. The Clinton surpluses would inevitably result in a downturn, just like every sustained budget surplus in U.S. history. Figure 7 shows the federal government balance as a percent of GDP (deficit or surplus) and periods of recession. (The sign of the government balance is reversed, so that a budget surplus is shown as a negative number.) Every recession since World War II was preceded by a government surplus or a declining deficit-to-GDP ratio, including the recession following the Clinton surpluses. Recovery from that recession resulted from renewed domestic private sector deficits, although growth was also fueled by government budget deficits that grew to 4 percent of GDP. However, as shown below, the Bush recovery caused tax revenues to grow so fast that the budget deficit fell through 2007, setting up the conditions for yet another economic collapse.

Just as surpluses precede recessions, large (nondiscretionary) budget deficits almost always result from recessions because of automatic stabilizers. When the economy slides into recession, tax revenues fall as economic activity declines. Social transfer payments, particularly unemployment benefits, increase automatically as more people lose their jobs. Despite all the conservative uproar against Obama's stimulus plan, the largest portion of the deficit increase to date has come from automatic stabilizers rather than from discretionary spending. This is observable in Figure 8, which shows the growth rate of tax revenues (mostly automatic, moving with the cycle because income and payroll taxes depend on economic performance), government consumption expenditures (somewhat discretionary), and transfer payments (largely automatic) relative to that in the same quarter a year earlier.

In 2005, tax revenues were growing at an accelerated rate of 15 percent per year—far above the GDP growth rate (hence, reducing nongovernment sector income) and above the government spending growth rate (5 percent). As shown in Figure 8, this fiscal tightening was followed by a downturn—that automatically slowed growth of tax revenue. While government consumption expenditures remained relatively stable during the downturn (after a short spike in 2007–08), the tax revenue growth rate dropped sharply from 5 percent to negative 10 percent within just three quarters (from the fourth quarter of 2007 to the second quarter of 2008), and to negative 15 percent by the first quarter of 2009. Transfer payments have been growing at an average quarterly rate of 10% (relative to the same quarter the previous year) since 2007. Decreasing taxes, coupled with increasing transfer payments, have automatically pushed the budget into a larger deficit, notwithstanding the flat consumption expenditures.

These automatic stabilizers, not the bailouts or stimulus package, are the reason why the U.S. economy has not been in a free fall comparable to that of the Great Depression. When the economy slowed, the budget automatically went into a deficit, placing a floor under aggregate demand. And in spite of all the calls to reign-in deficits, the truth is that deficits will not come down until the economy begins to recover. Even if we eliminated welfare payments, Medicaid, Medicare, military spending, earmarks, Social Security payments, and all programs except for entitlements; and also stopped the stimulus injections, shut down the education department, and doubled corporate taxes, the *New York Times* estimates that the budget deficit would still be over $400 billion. This example further demonstrates the nondiscretionary nature of the budget deficit. And, of course, this example doesn't consider how much more tax revenues would fall and transfer payments would rise if these cuts were actually

undertaken. With the current automatic stabilizers in place, the budget cannot be balanced, and attempts to do so will only damage the real economy as incomes and employment fall.

To summarize, fiscal policy should focus on results rather than follow conventional/ceremonial ideas about what is sound and what is not. A sovereign government spends by crediting bank accounts, while taxation debits them. Rather than "paying for" government spending, the real macroeconomic purpose of taxes is to keep private income and spendingat a noninflationary level. Clearly, in current circumstances, it would be a big mistake to raise taxes now—when the danger is unemployment and deflation.

Government budgets and self-imposed constraints

Guided by flawed economic thinking, governments worldwide have imposed constraints on their fiscal capacity to fully utilize their labor resources. Examples of self-imposed constraints include issuing bonds before government can spend, erecting barriers between fiscal and monetary authorities (and giving the latter more "independence"), and setting debt ceilings, deficit limits, and so on. A sovereign government doesn't need to sell bonds in order to spend because it can simply tell its central bank to credit its account by as much as it needs prior to writing checks on that account. Alternatively (and much more sensibly), the central bank and treasury can be consolidated so that the treasury can credit bank accounts along with its spending.

Even though governments have adopted a variety of self-imposed constraints, they are not normally binding. For example, the prohibition on the sale of treasury bonds directly to the central bank is easily circumvented. When the U.S. Treasury does not have sufficient funds in its checking deposit at the Fed, it sells bonds to special depositories (private banks) that are allowed to buy the bonds by crediting the Treasury's checking deposit held temporarily in private banks. The Treasury then transfers its deposit to the Fed before spending (it can only write checks on deposits at the Fed). This would normally result in a reserve debit from the bank accounts, but the Fed allows a "float" to postpone the debit because Treasury spending will restore the reserves. The final result is that the banks hold Treasuries and the customers have demand deposits. If the banks prefer to hold reserves, the Fed engages in an open market purchase—buying bonds and crediting bank reserves (as discussed above this is equivalent of moving funds from bank "saving accounts" to "checking accounts" at the Fed). The net effect is exactly the same as if the Fed had bought the bonds directly from the Treasury.

As another example, the Treasury must get the approval of Congress to expand its "borrowing limit" when it approaches its debt ceiling. After members of Congress dutifully wring their hands and declaim the burden placed on future generations by the administration, the debt limit is increased. In any case, bond sales are a completely voluntary and self-imposed operation for a sovereign government. As discussed, bonds are merely an interest-earning "saving account" alternative to low-earning reserves "checking account", and they are used by the Fed to hit its interest rate target. A central bank can simply pay interest on reserves (as Canada has done for a long time and the Fed is now doing) and the government can dispense entirely with selling bonds and worrying about debt ceilings. The Fed would then act as the Treasury's bank by taking a Treasury IOU and crediting Treasury's account when it wanted to spend. When the Treasury spent, the Fed would credit private banks with reserves and the banks would credit their customers' bank deposits. Taxes would reverse this procedure.

Under this procedure, budget deficits would generate reserve growth (bank deposits at the Fed that are the Fed's liability) that is offset by growth of the Treasury's liability to the Fed (the Fed's asset). Congress (or the Fed) could set the interest rate on the Treasury's liabilities held by the Fed that are used for accounting purposes. Since Fed earnings above a 6 percent return go directly to the Treasury, the Treasury in effect would pay most of the interest to itself. The rest would go to the Fed to help cover the costs of paying interest on reserves at the overnight rate chosen by the Fed (and distributing profits to its share-holding member banks). This would greatly simplify procedures, make the operations more transparent, and allow everyone to stop worrying about federal government debt. Since reserves are not counted as debt, there would be no publicly held debt. It should be recognized that Treasury IOUs held by the Fed simply represent internal accounting—the government owing itself and paying interest to itself. Any interest paid out by the Fed to banks holding reserves should be booked as government expenditure; that is, a subsidy to the banking system.

The rate paid today on reserves (and on short-term government bills) is a discretionary-policy variable. One of the huge fears about budget deficits is that the government might find that it would have to pay ruinous interest rates one day to get anyone to hold its debt. Let's presume that the federal government followed the proposal laid out in the previous paragraphs but the Fed decided to pay zero interest on reserves. With Treasury spending, the bank accounts of recipients would be credited with deposits and the banks' accounts at the Fed would be credited with nonearning reserves. Presumably, the banks would not want to hold any excess reserves (reserves above the amount required by the Fed or needed for clearing). They would offer reserves in the overnight Fed funds market, driving the rate to zero. Since the Fed would be paying zero on reserves, the "market equilibrium" rate would be zero. And try as they might, banks in the aggregate cannot get rid of reserves, since only tax payments or cash withdrawals reduce reserve holdings.

Would the banks refuse to allow their customers to receive deposits from the Treasury? No, since that would cause them to lose customers. It is possible that banks would begin to impose fees on deposit accounts to cover some of the costs of offering transaction accounts, while holding (nonearning) reserves in their portfolios. If that is not a desired outcome, the government could subsidize the private payments system by paying a positive interest rate on reserves—as discussed above, this simply means that the Fed credits bank "checking accounts", much as banks pay interest by crediting customers' checking accounts.

Let's return to the current system, in which the federal government issues bonds and the Fed pays a low interest rate on reserves. As discussed, deficit spending by the Treasury results in net credits to banking system reserves. If excess reserves are generated, banks offer them in the overnight market, which pushes the Fed funds interest rate down to the "support rate"—the rate paid on reserves. If the Fed prefers to maintain a higher Fed funds rate, it can engage in an open market sale of Treasuries and substitute them for reserves. This is how it maintains the Fed funds market rate at its target overnight rate—a spread above the rate it pays on reserves. If the Treasury only issues short-term bills, its interest rate will be determined by substitution in the overnight lending market. In other words, the rate on Treasury debt will be set relative to the Fed's overnight target rate. This result holds no matter how big the deficit or how much government debt is issued, so long as its maturity is short enough to be a close substitute for overnight interbank lending.

This means that the government doesn't need to allow the markets to determine the interest rate it pays on its debt. And even if Treasury chose to issue longer-term bonds, the Fed could actually set interest rates of different maturities if it were willing to deal in bonds of different maturities. Effectively, government could offer the equivalent of a range of

"certificates of deposit" with different maturities and interest rates—exactly what banks do with their certificates of deposit. If the government offered, say, 4 percent on "deposits" of 30 years but found no takers, that would be perfectly fine. It could either adjust the 30-year rate to try to find buyers—or, better, simply let buyers choose shorter maturities at lower rates.

This leads us back to the concern about foreign holders of debt. Foreign sellers of goods, services, or assets to the United States receive dollar credits, usually to a foreign branch of a U.S. bank or to a correspondent bank that receives a credit to its reserve account (or to the reserve account of a "mother" bank). If this bank preferred domestic currency reserves, the dollar reserves would end up in the account of its central bank. In any case, the holder of dollar reserves will probably try to find a higher interest rate—offering reserves in the overnight market or buying U.S. Treasuries. All of the analysis presented previously applies here, but with one wrinkle: the foreign holder could decide to exchange the dollar reserves for other currencies. Of course, the exchange cannot occur unless there is someone with the desired currency who is willing to exchange another currency for dollars. It is conceivable that, as portfolios of currency reserves were adjusted, exchange rates would adjust with the U.S. current account deficit placing downward pressure on the dollar.

While the conventional wisdom is that the Fed could keep the dollar from depreciating by raising domestic interest rates, there is plenty of empirical evidence to doubt the efficacy of interest rate adjustments impacting exchange rates. As argued above, the decision to sell products to the United States is not independent of the decision to accumulate foreign currency. We are skeptical that the interest rate paid on foreign currency reserves is as important as the decision to export or accumulate foreign currency. As discussed above, we see the U.S. current account deficit as the flip side of the coin to foreign desire to accumulate dollar assets. In the first instance, these claims take the form of reserves at the Fed. Holders will naturally try to earn the maximum return consistent with their appetite for risk, and hence prefer U.S. Treasuries that pay more than reserve deposits at the Fed. But they will take what they can get.

In conclusion, there is no financial constraint on the ability of a sovereign nation to deficit spend. This doesn't mean that there are no real resource constraints on government spending, but these constraints, not financial constraints, should be the real concern. If government spending pushes the economy beyond full capacity, then there is inflation. Inflation can also result before full employment if there are bottlenecks or if firms have monopoly pricing power. Government spending can also increase current account deficits, especially if the marginal propensity to import is high. This could affect exchange rates, which could generate pass-through inflation.

The alternative would be to use fiscal austerity and try to keep the economy sufficiently depressed in order to eliminate the pressure on prices or exchange rates. While we believe that this would be a mistake—the economic losses due to operating below full employment are almost certainly much higher than the losses due to inflation or currency depreciation—it is an entirely separate matter from financial constraints or insolvency, which are problems sovereign governments do not face. However, as discussed in the next section, nonsovereign governments do face financial constraints and can be pushed into involuntary default.

Countries with nonsovereign monetary systems

Recently, all eyes have been on Greece, which has been harshly criticized for fiscal irresponsibility—not to mention cooking its books and masking its debt and deficit levels with

the help of Goldman Sachs. With an estimated budget deficit of around 13 percent of GDP (more than quadruple the Maastricht criterion) and debt of 120 percent of GDP, Greek bonds have been downgraded by rating agencies. Even with already high interest rates on its debt, Greece is having a hard time finding investors willing to buy its bonds, and has asked the International Monetary Fund and European Union members to help by providing funding. While a bail-out package will be forthcoming, it comes with crushing fiscal austerity requirements. We do not believe this will ultimately be successful, and expect that the crisis will spread to other euro nations.

To intensify scare tactics, deficit hawks use Greece as an example of what awaits the United States if it doesn't tighten its fiscal belt. But in doing this, the hawks fail to distinguish between a nonsovereign (Greece) and sovereign (United States) country. We agree that the concern about Portugal, Ireland, Italy, Greece, and Spain (PIIGS) and other euro countries is justified. But considering the PIIGS as analogous to the United States is a result of the failure of deficit critics to understand the differences between the monetary arrangements of sovereign and nonsovereign nations. Greece is a user, not an issuer, of the euro. In that respect, it is more like California or even Orange County, both of which are users of the dollar. It is a serious mistake to argue that a sovereign nation like the United States should be constrained in the same manner as Greece or Orange County.

Eurozone countries have faced two types of problems. First, they have given up their monetary sovereignty by abandoning their national currencies and adopting a supranational one. And by divorcing fiscal and monetary authorities, they have relinquished their public sector's capacity to provide high levels of employment and output. Nonsovereign countries are limited in their ability to spend according to taxation and bond revenues, and this applies perfectly well to Greece, Portugal, and countries like Germany and France. No U.S. state has a budget deficit or debt ratio relative to GDP that comes close to that of Germany, let alone that of Greece or Italy, even with the devastating recession that is killing state budgets, yet they are already meeting market resistance to new borrowing precisely because they are nonsovereign. We suspect that eurozone countries have been allowed to exceed the limits imposed by markets on U.S. states because there is some market uncertainty about the nature of these countries. Are they nonsovereign? Will their neighbors rescue them? Will the European Central Bank (ECB) or International Monetary Fund (IMF) rescue individual nations? The answers seem more clear in the case of the U.S. states: neighboring states will not aid a defaulting state, no international lender is going to intervene, and a full federal government bailout is unlikely (while it is probable that there would be some sort of rescue, debt would probably face at least some haircut, and some holders might be excluded).

Second, the eurozone countries have agreed to abide by the Maastricht Treaty, which restricts budget deficits to only 3 percent of GDP and debt to 60 percent of GDP. Even if these countries are able to borrow and finance their deficit spending (e.g., Germany and France), they are bound not to use fiscal policy above those limits. In response, countries have resorted to different means in keeping their national economies afloat—fostering the export sector (in the case of Germany) or cooking books through Wall Street wizardry (Greece). These constraints have proven to be flexible, but that does not mean they do not matter. When a nation exceeds mandated limits, it faces punishment by European institutions and by markets. There is competition within the eurozone for ratings, with Germany usually winning and enjoying lower credit default swap pricing that allows it to issue euro debt at a lower interest rate. That in turn lowers its interest spending and deficits in a nice, virtuous cycle. Countries such as Greece that exceed the limits the most are punished with high interest rates that drive them into a vicious death spiral, with further credit downgrades and higher interest rates as deficits continue to rise.

Although the "Greek tragedy" should be a real concern, all of the proposed solutions share the same flaws that spring from a mistaken understanding of how public finance works in sovereign nations. Germany, France and the IMF have agreed to help Greece if it becomes more responsible in balancing its budget and retiring its debt (the details are not yet known). Greece is therefore forced to cut its budget deficit in a recession, which could worsen the eurozone's situation since it grew by only 0.1 percent in the fourth quarter of 2009. Greece will try to reduce its deficit by cutting public sector wages and pensions, a step that would further exacerbate the problem by decreasing incomes and employment. Indeed there is no guarantee that fiscal austerity will actually reduce the deficit—since slower growth will reduce tax revenue in another vicious cycle. As the eurozone stagnates, members such as Portugal, Italy, and Spain could face the same situation as Greece. And so it goes. It is important to realize that even Germany and France are threatened: Germany because it relies on exports to other eurozone members to keep up its employment, and France because its banks are major creditors of the PIIGS.

There are two real solutions for Greece and other eurozone members. First, members could exit the eurozone, regain monetary sovereignty, and run budget deficits that are large enough to achieve full employment. They would have to default on their euro-denominated debt because it would become even more difficult to service the debt in euros (especially if trade sanctions were slapped on the countries that leave). By doing so, individual countries would regain control of domestic policy space and spend like the United States—by crediting bank accounts. This option would relieve the newly sovereign governments from being at the mercy of markets, rating agencies, and other countries, and enable them to fully utilize their labor resources. There would be, however, transitional costs—including possible sanctions placed on them by other nations as well as political and market uncertainty.

The second and preferred solution to help all eurozone countries facing default is to create a supranational fiscal authority similar to the U.S. Treasury that is able to spend like a sovereign government. Alternatively, countries could be allowed to have overdrafts in their ECB accounts that enable them to spend euros like a sovereign government. Warren Mosler has proposed a viable stopgap measure whereby the ECB would create and distribute one trillion euros among members on a per capita basis so that each individual country could regain control over spending. This measure would give Euroland the time to come up with a more permanent solution, such as creating a supranational treasury that could spend as much as 10 or 15 percent of the region's GDP (the European Parliament's budget is currently less than 1 percent of GDP, which is far too small to generate a sufficient level of aggregate demand). Again, the distribution of spending could be decided by individual member states.

More generally, the failure to distinguish sovereign government debt from nonsovereign government debt and the debt of households and firms calls into question the results of another Reinhart and Rogoff study (2009b), which lumps together government and private debt, and argues that a private or public debt buildup poses systemic risks. While we agree that an excessive private debt buildup is unsustainable, the same cannot be said about sovereign government debt. It therefore makes no sense to add these debts together. Also, we need to clearly distinguish between foreign- and domestic-denominated debts. A sovereign government's debt denominated in its own currency cannot be subject to default risk nor can it cause slow growth, as it represents the nongovernment sector's net financial wealth. Many have claimed that the Reinhart and Rogoff studies (2009a,b) demonstrate that high debt ratios lead to slow economic growth. Yet, if sovereign government debt is a private sector asset it is highly implausible to argue that putting more wealth into the hands of the nongovernment sector will generate slow growth.

The Reinhart and Rogoff studies fail to adequately distinguish between countries operating with different monetary regimes, and this distinction must be made when having a meaningful discussion about government finances. For example, the analysis doesn't distinguish between sovereign countries (e.g., the United States and United Kingdom) and countries that have given up their monetary sovereignty (e.g., the eurozone). Moreover, many countries changed their monetary system over the period (literally, centuries) covered by the Reinhart and Rogoff study. In the United States, for example, one cannot compare the period before and after 1973 as if nothing had changed. And for many countries, the dataset goes back to the early 19th century, when they were still on the gold standard and, hence, not sovereign (in our definition). The Reinhart and Rogoff study may apply to the United Kingdom before the 1930s (when it was still on the gold standard), but it doesn't apply today. Therefore, the finding that debt ratios above 90 percent of GDP are correlated with lower growth is not applicable to sovereign nations, since it seems to be driven by aggregating countries on a gold standard (or similar fixed exchange rate) with those that are sovereign (and issue their own floating-rate currency). Frankly, given the obvious confusion and conflation over different types of debt (sovereign government, nonsovereign government, and nongovernment) in the book, we cannot find any conclusions that are relevant to the current U.S. situation.

As explained earlier, the U.S. federal government budget moves countercyclically, so that low growth causes the budget deficit to expand. No doubt this response explains some of the correlation reported by Reinhart and Rogoff—that high debt ratios are associated with lower economic growth—but the causation is reversed for a sovereign nation, with slow growth causing deficits and raising debt ratios. In the case of a nonsovereign government, large deficits probably cause slow growth due to the imposition of austerity policies that are normally required for nations operating fixed exchange rates. This is the disadvantage of operating without a sovereign currency: both policymakers as well as markets will impose high interest rates on nonsovereign debt, and policymakers will probably try to raise taxes and cut spending to protect the currency peg. These policies lower growth but increase budget deficits (due to high interest rates and low growth) and generate the empirical correlation found in the study.

Further, it is conceivable that an expansion fueled by private sector debt will be followed by a period of low growth when private spending is depressed, since households and firms try to reduce debt ratios through increased savings. Given all of these complexities, the finding that debt ratios above 90 percent of GDP are correlated with lower economic growth provides no guidance for policymakers, especially those in sovereign nations.

Moreover, the 90 percent ratio is rather an arbitrary number. The debt thresholds selected by Reinhart and Rogoff (2009a) are based on their "interpretation of much of the literature and policy discussion." So far, however, no economist or policymaker has been able to come up with a nonarbitrary number for the debt-to-GDP ratio that has some economic meaning. The reason is that there is no such magic number applicable to all countries at all times. As discussed above, the government deficit (and debt) has to equal the balance of the private sector, which is based on its preference to save and import. It therefore varies among countries and between time periods.

Conclusion: The role of ideology

The hysteria about government deficits comes from a flawed understanding of how the monetary system works. It is questionable how much of this is ideological and how much is really a misunderstanding. Sovereign governments are led to believe that they need to

issue bonds and collect taxes to finance government spending, and that good policies should be judged in terms of fiscal austerity. Mainstream economics has guided policymakers to impose artificial constraints on fiscal policy and government finances, such as issuing bonds in response to deficits, setting debt ceilings, forbidding the central bank to buy treasury debt directly, and allowing markets to set interest rates on government bonds. To further dupe the public, a strong case is made for independent monetary policy and separating monetary authority from fiscal authority in order to reduce the influence of political pressures. All of these constraints are self-imposed and voluntary.

Ideologically motivated economists have praised the merits of monetary policy in controlling inflation by declaring that price stability is all that is necessary to stabilize the economy. They leave little room for stabilizing fiscal policy in their models. They warn the public that government spending causes inflation, and that if budget deficits are not controlled, we could become the next Zimbabwe (the most recent example of hyperinflation) or the Weimar Republic (Germany in the 1930s). The historical context and case specifics are ignored, while the presumption that such analogies to failed states or household budget constraints is sufficient. Proposals supporting deficit spending as a means of dealing with economic crises are met with warnings that government debt will burden future generations with high taxes. This implies that it is better to pay for our excesses now than to pass along our problems to our grandchildren. Moreover, concerns about government deficits and debts have masked the real issue: deficit hawks are unwilling to allow a (democratic) government to work for the good of the people.

We accept that there are real differences of opinion regarding the proper role of government in the economy. Some would like to see the functions of government curtailed; others would like to see them expanded. These are legitimate political stances. What is not legitimate is to use fear over deficits to restrain government from achieving the public purpose that is democratically approved. A debate that is freed from the constraints imposed by myths about how government really spends would allow us to move forward to gain consensus on the public purpose the American people expect government to pursue.

Notes

1. This is the total outstanding debt ratio. The relevant figure is the portion held by the public, which reaches only 73 percent.

2. We will not revisit the wisdom of such a scheme but merely argue that for all intents and purposes, Social Security's Treasury holdings really amount to internal record keeping—a reminder that Social Security, taken alone, runs surpluses now and that the Treasury will have to cover Social Security's shortfall someday. Yet that has nothing to do with the overall budget stance of the federal government—which can be balanced or in surplus or deficit regardless of the balance of individual federal government programs.

3. There is the belief that the debt owned by Social Security should be counted because it reflects a future obligation of government to future beneficiaries. However, the government is obliged meet those obligations whether or not Social Security owns Treasuries, and it will meet its obligations in exactly the same manner whether or not it holds Treasuries (see Wray 2005).

4. This can also be looked at in terms of the leakage-injection approach: budget deficits as well as domestic private deficits are injections that must equal the leakage of current account deficits. Given the propensity for net imports and the federal government's surplus, the domestic private sector's deficit must be that much larger to match the leakages due to current account deficits plus the government surplus.

References

Boskin, M. 2010. "When Deficits Become Dangerous." *The Wall Street Journal*, February 11. *Economic Report of the President*. 2010. Washington, D.C.: U.S. Government Printing Office. February.

Mankiw, G. 2010. "What's Sustainable about This Budget?" *The New York Times*, February 13.

Rampell, C. 2010. "Stiffening Political Backbones for Fiscal Discipline." *The New York Times*, February 12.

Reinhart , C. and K. Rogoff, 2009a." Growth in a Time of Debt." Draft paper prepared for the *American Economic Review Papers and Proceedings.* December 31.

———. 2009b. *This Time Is Different: Eight Centuries of Financial Folly*. Princeton, N.J.: Princeton University Press.

Wray, L. R. 2005. "Social Security's 70th Anniversary: Surviving 20 Years of Reform." Policy Note 2005/6. Annandale-on-Hudson, N.Y.: Levy Economics Institute of Bard College. September.

SUGGESTED CITATION:
Yeva Nersisyan and L. Randall Wray, "Deficit hysteria redux? Why we should stop worrying about U.S. government deficits", *real-world economics review*, issue no. 53, 26 June 2010, pp. 109-128,
http://www.paecon.net/PAEReview/issue53/NersisyanWray53.pdf

The social cost of carbon[1]

Frank Ackerman and Elizabeth A. Stanton [Stockholm Environment Institute, USA]

The social cost of carbon may be the most important number you've never heard of. U.S. climate legislation may or may not make it through Congress this year, but in the meantime, the Environmental Protection Agency is moving ahead, authorized by the Supreme Court to limit greenhouse gas emissions. The Department of Energy is setting energy efficiency standards for residential appliances and commercial equipment, based in part on their contribution to climate change. Other agencies may address the same issues, when their regulations affect energy use and carbon emissions.

The social cost of carbon (SCC), defined as the estimated price of the damages caused by each additional ton of carbon dioxide (CO_2) released into the atmosphere, is the volume dial on government regulations affecting greenhouse gases: The higher the SCC is set, the more stringent the regulatory standards. This white paper explains how economists estimate the social cost of carbon, why the Obama Administration's current analyses are on a path to grossly underestimating it, and why relying on the SCC in the first place may be unproductive.

The EPA, DOE, and other agencies are deciding on values to assign to the SCC in the next few months as part of "rulemaking" processes that are couched in very technical terminology and largely invisible to the general public. In theory, it appears possible to derive the SCC from economic analysis, and the administration appears to have done so. In reality, it's not so simple: Any estimate of the SCC rests on a number of value judgments and predictions about uncertain future events, and so far, the administration has made choices that lead to very low SCC values. In an interim and then a revised analysis, an interagency working group has presented multiple scenarios and possible values for the SCC; the interim analysis suggests, and the revised analysis explicitly endorses, a "central" estimate of $21 per ton of CO_2 in 2010. This amounts to roughly 20 cents per gallon of gasoline, an extremely modest price incentive for carbon reduction. If adopted, this obscure number will have immense practical consequences: A low SCC could result in ineffectual regulations that lead to few if any reductions in U.S. emissions until Congress passes a climate bill.

Even greater harm could result if Congress interprets the $21 SCC as an endorsement of that level for a carbon tax or permit price. This could clash with the widely discussed, science-based goal of achieving an 80 percent reduction in U.S. emissions by 2050, an objective that will almost certainly require a much higher price on carbon. In the revised analysis, the central SCC estimate rises only to $45 per ton (in 2007 dollars) by 2050.[2] If climate economics is (mistakenly, in our view) interpreted as supporting an SCC of only $21 today and $45 by mid-century, it could also be interpreted as advocating only the emission reductions that would result from those prices. That is, working backwards from the proposed SCC, one could infer that the appropriate cap on carbon emissions is much weaker than those found in recent legislative proposals. The resolution to this paradox is that, as we argue in this paper, the $21 SCC is based on flimsy analyses and multiple mistakes. Sound

[1] A Report for the Economics for Equity and the Environment Network, www.e3network.org.

[2] U.S. Department of Energy (2010), "Final Rule Technical Support Document (TSD): Energy Efficiency Program for Commercial and Industrial Equipment: Small Electric Motors," Appendix 15A (by the Interagency Working Group on Social Cost of Carbon): "Social Cost of Carbon for Regulatory Impact Analysis Under Executive Order 12866," available online at http://www1.eere.energy.gov/buildings/appliance_standards/commercial/sem_finalrule_tsd.html.

economic analysis would show that the SCC should be much higher, and thus could be consistent with the carbon prices required to achieve science-based targets for emission reduction.

Calculating the SCC is a new undertaking for the administration, and these initial estimates may represent work in progress rather than a final answer. In its first attempts, however, the administration's interagency working group has left itself plenty of room for improvement.

1. The Back Story

A ton of CO_2 is the basic unit of emissions for climate policy, but it may be hard to visualize – especially since it's a colorless, odorless gas that mixes into the air around us. In the United States, one ton of CO_2 is emitted, on average, by:

- A family car every two and half months.[3]
- A household's use of heating and cooking fuel every four months (if energy use were spread equally throughout the year). That's every four years in Hawaii or every six weeks in Maine.[4]
- A household's use of electricity every six weeks.[5]
- The typical use of a microwave oven every seven years or of a refrigerator every 15 months.[6]

U.S. residents emitted 21 tons of CO_2 per person in 2005: 33 percent from transportation, 15 percent from residential electricity, 6 percent from home heating and cooking, and the remaining 46 percent from industry, retail stores, and government.[7]

Each person's annual 21 tons of CO_2 add to the stockpile of greenhouse gases in the atmosphere. The more CO_2, the hotter the average global temperature (the "greenhouse effect"), the faster sea levels rise (warmer waters expand to take up more room, while glaciers and polar ice caps melt), and the more our weather patterns diverge from historical trends (changes to rainfall, more intense storms).

How fast are we making the climate worse? The amount of CO_2 in the air was 280 parts per million (ppm) before the industrial revolution, in 1750, and has now reached 385

[3] Average U.S. passenger fuel efficiency for 2007 was 22.5 mile per gallon (BTS RITA Table 4-23, http://www.bts.gov/publications/national_transportation_statistics/html/table_04_23.html). Motor gasoline emissions coefficient, 19.564 lbs. per gallon (USEIA, Voluntary Reporting of Greenhouse Gases Program, http://www.eia.doe.gov/oiaf/1605/coefficients.html). U.S. miles per passenger car in 2001, 12,000 (USEIA NHTS Table A3, http://www.eia.doe.gov/emeu/rtecs/nhts_survey/2001/tablefiles/table-a03.pdf).

[4] For direct residential emissions calculations see Stanton, E.A., F. Ackerman, and K. Sheeran (2009), *Greenhouse Gases and the American Lifestyle: Understanding Interstate Differences in Emissions*. Stockholm Environment Institute, Economics for Equity and the Environment Network, available online at http://www.e3network.org/papers/NRDC_state_emissions_report.pdf . Data used here are updated to 2005. Number of households in 2005 by state, ACS 2005 B11001, http://www.census.gov/.

[5] For electricity emissions calculations see Stanton, Ackerman and Sheeran (2009). Data used here are updated to 2005. Number of U.S. households in 2005, 111 million (ACS 2005 B11001, http://www.census.gov/).

[6] Ibid. Average energy use for appliances, 200 kWh/year for microwaves, 1100 kWh/year for refrigerators (USDOE Web site, http://www1.eere.energy.gov/consumer/tips/appliances.html).

[7] For methodology and data sources see Stanton, Ackerman and Sheeran (2009). Data used here are updated to 2005.

ppm. Doubling the concentration of CO_2 – on current trends we will reach 560 ppm, double the pre-industrial level, within this century – has been widely expected to increase the average global temperature by 3°C (5.4°F),[8] but recent research has called this into question. Newer studies are suggesting that doubling the atmospheric CO_2 concentration could raise that average temperature by as much as 6°C (11°F).[9] The size of the temperature increase associated with a doubling of atmospheric CO_2, a number referred to as the "climate sensitivity," is crucial to the scientific analysis of climate change.[10]

The purpose of emission reductions is to limit the change in average global temperature and related climate conditions; many scientists believe that any warming beyond 2°C (3.6°F) would put the world at too high a risk of catastrophic, irreversible consequences.[11] Already, CO_2 concentrations are well above pre-industrial levels, and CO_2, once emitted, stays in the atmosphere for a long time. This means that even if we could immediately stop all greenhouse gas emissions, there would still be a gradual temperature increase over the next century. The more we can slow down that increase, the easier it will be for human societies to adapt with careful planning and new technologies. Every ton of CO_2 that we can keep out of the atmosphere slows climate change, helps to hold temperatures under that 2°C threshold, and reduces the risk of the worst kinds of damage.

But reducing emissions also carries a cost – including the price of new "green" energy technologies, and more efficient appliances, vehicles, and heating and cooling systems. The policies used to reach this goal may leave households facing bigger energy bills. So to help determine how aggressively to act to cut emissions, policymakers weigh those costs against the cost of inaction, or of less-aggressive action. That's where the social cost of carbon comes in: It asks, how much will each ton of CO_2 that we release into the atmosphere cost us in damages, both today and in the future? If the answer is a big number, then we ought to make great efforts to reduce greenhouse gas emissions. If it's a small number, then the case for reduction is weaker, and only easy or inexpensive changes seem warranted, at least in narrowly economic terms.

For example, if the SCC had a value of $5 of present and future damages per ton of CO_2, we would be willing to pay up to $5 to prevent a ton from being released (just as you would put $4 in a parking meter to avoid a $5 ticket). But if the cost of a particular measure to reduce emissions had a higher price tag than $5 per ton, we might instead accept those future damages (just as you would prefer the $5 ticket to putting $6 in the meter).

This is why the SCC is so important: The policy choices the government makes would be very different if it estimates climate damages not at $5 but at, say, $500 per ton of carbon

[8] Intergovernmental Panel on Climate Change (2007), *Fourth Assessment Report: Climate Change 2007* (AR4).

[9] Hansen, J. et al. (2008), "Target Atmospheric CO_2: Where Should Humanity Aim?" *The Open Atmospheric Science Journal* 2: 217-231, and IPCC (2007).

[10] For recent analyses highlighting scientific concern about climate sensitivity see Roe, G. H., and M. B. Baker (2007), "Why is Climate Sensitivity So Unpredictable?" *Science* 318: 629-632; Clement, A. C., R. Burgman, and J. R. Norris (2009), "Observational and Model Evidence for Positive Low-Level Cloud Feedback," *Science* 235: 460-464; Solomon, S. G.-K. Plattner, R. Knuttic, and P. Friedlingsteind (2009), "Irreversible climate change due to carbon dioxide emissions," *PNAS* 106(6): 1704-1709; and Schellnhuber, H. J. (2008), "Global warming: Stop worrying, start panicking?" *PNAS* 105(38): 14239–14240.

[11] See Ackerman, F., E.A. Stanton, S.J. DeCanio, E. Goodstein, R.B. Howarth, R.B. Norgaard, C.S. Norman, and K. Sheeran (2009), *The Economics of 350: The Benefits and Costs of Climate Stabilization*, Stockholm Environment Institute, Economics for Equity and the Environment Network, available online at http://www.e3network.org/papers/Economics_of_350.pdf , and IPCC (2007).

(in the same way that a $500 parking fine would make you pay much more attention to putting money in the meter). Right now, of course, the price of carbon emissions is zero.

2. Uses of the Social Cost of Carbon

All current proposals for climate policy are based on the price of carbon emissions, whether it's through a carbon tax, market allowances, or through regulation by government agencies.

Carbon tax: Under this option – which is popular with some economists, but anathema in actual policy debates – the price per ton of carbon is applied as a tax on fuels. This can be done either at the well-head and the border, or at the point of consumption (a gasoline tax, for example). The government collects the taxes and can use the revenue for virtually any purpose: to reduce other taxes, to invest in clean energy, to assist workers transferring out of the most polluting industries, and so on.

Emission allowance markets: In a "cap and trade" scheme, a limited number of carbon allowances are issued, and a secondary market forms to buy and sell the permits. In a "cap and dividend" system, carbon allowances would be auctioned off, with the revenue returned to the public. Either way, a market is formed (the secondary market or the government auction) that sets the price of carbon through a give-and-take between buyers and sellers.

These two types of policies are symmetrical: A carbon price results in a reduction to emissions; a cap (or limit to emissions) results in a carbon price set by the market. If $X carbon price results in Y tons of carbon emitted, then a cap of Y tons should result in exactly the same $X carbon price. However, the distributional consequences – who ends up with the tax or allowance revenue in their pockets – depend on the exact provisions of a particular climate policy.

Government regulation: A government agency such as the EPA or DOE can ban polluting technologies, require a set of green technologies, or impose performance standards such as emissions limits. Such regulations can be established with little or no reference to economic analysis, in the classic "command and control" mode; or they can be guided by cost-benefit calculations. Under the latter approach, a policy is approved if its cost (per ton of CO_2 eliminated) is less than the carbon price; and a policy is rejected as uneconomical if its per-ton cost is more than the carbon price.[12] The current analyses of the social cost of carbon will be used to apply this kind of logic to U.S. regulatory proposals.

The administration plans to set the carbon price by using data and analyses taken from current climate economics literature. Their method sounds simple: collect a variety of social cost of carbon estimates from the literature, tweak them for comparability, and use the resulting range of values in decision-making. The next section discusses the numerous problems with the administration's initial attempts at picking numbers for the SCC.

[12] For a deeper discussion of the process of calculating the social cost of carbon, see Stern, N. (2006), *The Stern Review: The Economics of Climate Change*, London: HM Treasury, Chapter 2, available online at http://www.hm-treasury.gov.uk/stern_review_report.htm: and Clarkson, R., and K. Deyes (2002), *Estimating the Social Cost of Carbon Emissions,* U.K. Department for Environment Food and Rural Affairs (DEFRA), Working Paper 140.

3. The Obama Administration and the Price of Carbon

The federal government's estimates of the SCC have been developed by the Interagency Working Group on Social Cost of Carbon, with participation from the Council of Economic Advisers, Council on Environmental Quality, Department of Agriculture, Department of Commerce, Department of Energy, Department of Transportation, Environmental Protection Agency, National Economic Council, Office of Energy and Climate Change, Office of Management and Budget, Office of Science and Technology Policy, and Department of the Treasury. The working group's interim estimates were used in DOE's final rule on energy efficiency in refrigerated vending machines[13] in August 2009, and in EPA's proposed rule on tailpipe emission standards for cars and light trucks[14] in September 2009. The working group's revised, current estimates were used in DOE's final rule on energy efficiency in small electric motors[15] in March 2010, and are expected to be incorporated into the final version of the tailpipe emission standard.

The working group's interim and revised analyses of the SCC have several features in common. Both rely heavily on averages of estimates from three climate economics models: DICE, FUND, and PAGE. Both experiment with a range of discount rates for valuing future outcomes (explained below), showing how the estimated SCC depends on assumptions about discounting.

The interim SCC analysis is the simpler of the two.[16] It follows the latest academic publications that present the three models, modifying them only to use differing discount rates. The revised analysis starts from the same point, performs a similar analysis of discount rates, and then goes on to modify the three models to consider a range of possible values for climate sensitivity, and to constrain them to match socioeconomic and emissions scenarios developed in another modeling exercise, the Energy Modeling Forum 22 (EMF-22) studies.[17]

We believe that the working group's interim and revised SCC estimates rely on a biased and incomplete reading of the economic literature on climate change. The methods used to set these values reveal an unexplained confidence in a handful of authors and models, and offer arbitrary, unsupported judgments as grounds for ignoring important alternatives. Most of the errors, omissions, and arbitrary judgments tend to reduce the estimate of the SCC; a corrected version of the same calculations, therefore, would likely result in a larger SCC – and more stringent regulations of greenhouse gas emissions.

---

[13] U.S. Department of Energy, Energy Conservation Program (2009), "Energy Conservation Standards for Refrigerated Bottled or Canned Beverage Vending Machines; Final Rule," 10 CFR Part 431, Federal Register vol. 74, no. 167, Aug. 31, 2009, pages 49914-44968, available online at http://www2.eere.energy.gov/buildings/appliance_standards/commercial/pdfs/bvm_final_rule_notice.pdf.

[14] EPA (2009), "Proposed Rulemaking To Establish Light-Duty Vehicle Greenhouse Gas Emission Standards and Corporate Average Fuel Economy Standards," EPA Docket EPA-HQ-OAR-2009-0472, Federal Register vol. 74, no. 186, Sept. 28, 2009, pages 49411-49418, available online at http://www.epa.gov/fedrgstr/EPA-AIR/2009/September/Day-28/a22516d.htm.

[15] U.S. Department of Energy (2010), Appendix 15A (see full reference above)

[16] For a detailed discussion of the interim analysis, see Ackerman, F. (2009), "Comments on EPA and NHTSA 'Proposed Rulemaking to Establish Light-Duty Vehicle Greenhouse Gas Emission Standards and Corporate Average Fuel Economy Standards,' " EPA Docket EPA-HQ-OAR-2009-0472, Federal Register vol. 74, no. 186, Sept. 28, 2009, pages 49454-49789, available online at http://www.sei-us.org/climate-and-energy/Ackerman_Sept2009Comments_on_EPA_GHG.pdf.

[17] The EMF-22 studies were published in a special issue of *Energy Economics* in 2009, available online at http://emf.stanford.edu.

Ethical judgments and omitted values imply that any SCC is incomplete

Some of the serious anticipated damages from climate change, such as loss of endangered species, cannot be quantified or monetized. Much of the climate economics literature used to inform the working group's estimates omits these values entirely, effectively giving them a value of zero. As a result, estimates of the SCC may be too low or logically incomplete, in the sense that they exclude crucial, unmonetized dimensions of climate damages.

Ethical judgments about the treatment of unmonetized damages play a role in any climate policy, complementing the quantitative calculations embodied in the SCC: What importance should be given to, for instance, the loss of endangered species, unique habitats and environments, and human lives and communities? Attempts to assign dollar costs to these priceless values leads to meaningless or offensive valuations (some of which are discussed below). Exclusion of them, however (or banishing them to the netherworld of "caveats" and verbal qualifications that are ignored in practice) amounts to treating them as being known to have no value at all. Ethical questions arise, as well, within the calculation of the SCC, particularly in the treatment of costs and benefits to future generations, a topic we address below.

The arbitrary choice of three models biases the analysis

The economic assumptions leading to the choice of the three models, DICE, FUND, and PAGE, are discussed at some length in the interim analysis. (The revised analysis simply says that these are three widely used models.) The interim analysis first takes Richard Tol's 2008 meta-analysis of estimates of the SCC as a starting point; attention is then restricted to peer-reviewed studies; three specific integrated assessment models – FUND, PAGE, and DICE – are selected, while others are ignored; and an unstated corollary is that the data sets developed by the authors of these three models are adopted without discussion. Each step of this process introduces arbitrary biases into the SCC estimate.

First, Tol's meta-analysis of SCC estimates, which describes itself as a comprehensive review of published research, is in fact a highly personal view of the economics literature, with a strong emphasis on Tol's own work. [18] It includes 211 estimates of the SCC, of which 112 come from Tol.[19] Disproportionate numbers also come from a few other authors and models. Every version of William Nordhaus' DICE model is included, despite the fact that the newer versions were created to update and replace the older versions.

Tol has not published 112 separate studies of the SCC; rather, he has counted multiple scenarios and sensitivity analyses within his own studies as separate estimates. He has extended the same treatment to some, but not all, other economists. For example, the Stern Review,[20] which included multiple scenarios and sensitivity analyses, is treated as only generating a single estimate of the SCC in Tol's meta-analysis. Thus the use of Tol's meta-analysis as a starting point is not a neutral decision; it introduces biases in favor of the work of Tol and Nordhaus, and against the Stern Review, among others.

[18] Tol, R., "The Social Cost of Carbon: Trends, Outliers and Catastrophes," *Economics* (e-journal), Vol. 2, 2008.

[19] Ibid, Table 2, for author counts.

[20] Stern (2006).

Second, insisting on peer review as an absolute criterion for inclusion in the SCC process also creates a bias. Indeed, a principal effect is to rule out consideration of the widely discussed Stern Review, which offered an innovative, rigorous analysis leading to a relatively high estimate of the SCC, $85 per ton of CO_2. Tol and some other economists have criticized the Stern Review for appearing as a government policy report rather than in a peer-reviewed journal. The level of professional review and detailed scrutiny applied to the Stern Review both before and after its publication was, however, far beyond the normal peer review process for articles published in academic journals. Following the publication of the Stern Review, the American Economics Association published a symposium on it in the *Journal of Economic Literature*, and invited Stern to give the prestigious Ely Lecture at the AEA's annual meeting in 2008; that lecture was published in the *American Economic Review*, the highest-status journal in the field.[21]

Third, the FUND, PAGE, and DICE climate economics models are not the only relevant climate economics models. The interim SCC analysis simply asserts without any documentation or other justification that "the FUND, PAGE, and DICE models now stand as the most comprehensive and reliable efforts to measure the economic damages from climate change."[22]

No evidence is offered to support that judgment; the reader must take it or leave it, on the personal authority of the authors of the proposed rule. The judgment, however, is not universal. The EPA's own "Climate Economic Modeling" Web page[23] makes no mention of FUND, PAGE, or DICE, but describes the ADAGE and MiniCAM models, among others. The three chosen models, misidentified as the "most comprehensive and reliable," are in fact among the simplest of all IAMs in current use.[24]

Finally, the data sets developed for FUND, PAGE, and DICE are not the only data that should be considered. The transparency of simple models like these allows a relatively clear view of the data and relationships that drive the model results. For climate economics models in general, including FUND, PAGE, and DICE in particular, the software and model relationships are often less decisive than the data inputs in shaping the results. Extensive experiments with DICE by a range of researchers have shown that with small, reasonable changes to the basic data, DICE can yield very different projections (our own contribution to that "modified DICE" literature is cited below). The procedure suggested in the tailpipe emissions case not only endorses three specific models; it implicitly endorses the data sets offered by the models' authors. Those data sets embody a number of controversial judgments.

FUND, originally developed by Richard Tol, relies on data from numerous studies of particular climate-related costs and impacts by Tol and his coauthors. In the problematic area of monetary valuation of the loss of human life, Tol argues that the value of life in a country depends on how rich it is: As he and two coauthors wrote in a paper on valuation of health, "Following Tol (2002a), we value a premature death at 200 times per capita income [i.e.,

[21] Stern, N., "The Economics of Climate Change," *American Economic Review* (2008), 98:2, 1-37.

[22] EPA (2009), "Proposed Rulemaking" (see full reference above).

[23] http://www.epa.gov/climate/climatechange/economics/modeling.html

[24] Stanton, E.A., F. Ackerman, and S. Kartha, "Inside the integrated assessment models: Four issues in climate economics," *Climate and Development* (2009), 1: 166-184.

average per capita income for the region where the death occurs]."[25] The assumption that higher-income lives are of greater monetary value than lower-income ones is morally offensive.

Tol and his coauthors also conclude, based on a series of mistakes and miscalculations, that the early stages of global warming will cause a huge *reduction* in mortality.[26] Valuing these allegedly saved lives at 200 times their per capita incomes creates a huge, spurious benefit of moderate warming, thereby reducing the net cost of climate damages and the SCC. The multiple mistakes in Tol et al.'s calculation of mortality reduction are explained in our response in the same journal.[27]

Chris Hope, the developer of PAGE, has responded to several objections to particular data inputs by converting them to uncertain parameters, allowing them to vary across a range of different values and looking at the average result. PAGE has produced many different estimates, including the Stern Review results which the interagency working group ignored, as well as the lower SCC values which the working group adopted. In a collaboration between Hope's research group and ours, we came to question PAGE's low projections of climate damages to the United States, even in the Stern Review version of the model.[28] The PAGE data set assumes that developed countries can and do engage in nearly costless adaptation to most climate damages in the next century. In addition, PAGE sets a relatively high temperature threshold for the onset of catastrophic damages, which seems inconsistent with recent scientific discussion of climate risk. Based on changes to these and other assumptions, we worked with Hope to produce several alternate estimates for U.S. and global damages due to climate change, ranging up to five to six times the PAGE defaults used by the working group.

The DICE model, developed by William Nordhaus, is known for its finding that the optimal climate policy is a very gradual one, starting on a small scale and expanding at a leisurely pace; Nordhaus refers to this as the "climate policy ramp." The gradualism of the default DICE projections is driven by the DICE estimate of climate damages, which is surprisingly low. One factor holding down the overall damage estimates is the assumed large benefit of warmer temperatures. On very thin evidence, Nordhaus assumes that most people in the world would be willing to pay for a warmer climate; he concludes that the optimal temperature is far above the current global average.[29] In work in progress, University of California-Berkeley economist Michael Hanemann has used up-to-date information to re-estimate each of the economic impacts of climate change included in the DICE damage

[25] Bosello, F., R. Roson, and R. Tol, "Economy-wide estimates of the implications of climate change: Human health," *Ecological Economics* (2006), 58: 579-591; quote from 585.

[26] Bosello, Roson, and Tol 2006; see Table 1, page 582, for projected changes in the number of deaths.

[27] Ackerman, F. and E.A. Stanton, "A comment on 'Economy-wide estimates of the implications of climate change: Human health,' " *Ecological Economics* (2008), 66:8-13.

[28] Ackerman, F., E.A. Stanton, C. Hope, and S. Alberth, "Did the Stern Review underestimate U.S. and global climate damages?" *Energy Policy* (2009), 37:2717-2721.

[29] Nordhaus, W., and J. Boyer, *Warming the World: Economic Models of Global Warming* (MIT Press, 2000), 84-85. The assumed positive value of warmer temperatures for most of the world is still visible in the "lab notes" documenting the data set for the newest version of DICE, http://nordhaus.econ.yale.edu/Accom_Notes_100507.pdf (page 24, "time use" column). For a critique, see Ackerman, F., and I. Finlayson, "The economics of inaction on climate change: A sensitivity analysis," *Climate Policy* (2006), 6: 509-526.

function, concluding that damages in the United States could be four times as large as the estimates implied by the DICE defaults.[30]

The conclusion is clear: The decision to rely exclusively on the FUND, PAGE, and DICE models and their underlying data sets imposes a narrow, biased filter on the economic analysis of the SCC. If agencies rely on these model results, especially in the absence of other analyses, they will almost certainly underestimate the social cost of carbon.

Casual, undocumented estimates are used to justify the choice of discount rate

Estimates of the social cost of carbon combine present and future damages together as one value, the total impact of an additional ton of CO_2. The process for combining costs from different time periods is called "discounting." The farther into the future that costs take place, the less these costs are assumed to matter in today's decision-making. But discounting also involves a judgment call: Future values can be discounted a lot, so that they have little bearing on our decisions; not at all, so they weigh equally with present costs, or somewhere in between. The higher the "discount rate" that is chosen, the less future costs are valued in present-day terms.[31]

When discounting is used to combine values from a short span of years, a market rate of interest is often taken to be an appropriate discount rate; this may be 5 percent or even higher. In theory, if we knew that climate damages would cost $100 ten years from now, we could invest $64 today at 5 percent interest to cover those costs in the future. To put this another way, at a 5 percent discount rate, a $100 cost ten years from now can be valued at $64 today; anyone who expects to incur a $100 cost ten years from now could put $64 in the bank in 2010, and withdraw $100 in 2020.

However, when discounting takes place across a longer span of time, the logic of using market rates becomes muddied. Climate policy is inescapably concerned with mitigation costs incurred today that will have their greatest benefits a century or more into the future, yet there is no single individual who can compare her own costs today with benefits 100 years from now. The choice of a discount rate for intergenerational impacts is an ethical judgment, not a data point that can be found in the financial pages. Lower discount rates, decreasing rates over time, and even a zero discount rate (no discounting) can be used to show that our society takes seriously the costs to be suffered by future generations.[32]

The interim analysis recommends two alternate discount rates, 3 percent and 5 percent, for use in calculating the SCC, while noting that "decisions based on cost-benefit analysis with high discount rates might harm future generations."[33] Casual estimates and

[30] Hanemann, W.M. (2009),"What is the Economic Cost of Climate Change?" University of California-Berkeley. On errors in an influential early analysis of agriculture and climate change coauthored by Nordhaus, see Schlenker, W., W.M. Hanemann and A.C. Fisher, "Will U.S. Agriculture Really Benefit from Global Warming? Accounting for Irrigation in the Hedonic Approach," *American Economic Review* (March 2005) 395-406.

[31] For a detailed discussion of discounting, see Stern (2006) and Arrow, K. J., W.R. Cline, K.-G. Maler, M. Munasinghe, R. Squitieri, and J.E. Stiglitz (1996), "Chapter 4 - Intertemporal Equity, Discounting, and Economic Efficiency," in *Climate Change 1995 - Economic and Social Dimensions of Climate Change, Contribution of Working Group III to the Second Assessment Report of the IPCC,* J. P. Bruce, H. Lee and E. F. Haites, Eds. New York, NY: IPCC and Cambridge University Press: 125-144.

[32] For a more detailed discussion see Ackerman (2009), "Comments on EPA and NHTSA 'Proposed Rulemaking.' "

[33] EPA (2009), "Proposed Rulemaking" (see full reference above).

unsupported judgments are used to justify discount rates that are inappropriately high for an analysis that spans several generations. The Office of Management and Budget guidelines encourage sensitivity analysis with discount rates below 3 percent for intergenerational problems.[34] The revised SCC analysis takes a timid step in that direction, adding a discount rate of 2.5 percent, along with 3 percent and 5 percent.

Catastrophic climate risk is left out of the calculations

The administration's estimates of the social cost of carbon largely omit the risk of catastrophic climate damage. (DICE includes the expected value of a moderately catastrophic economic downturn, with a magnitude based on a very old opinion survey; PAGE includes a Monte Carlo analysis of the risk of a similar-sized catastrophe; FUND ignores the issue.) The interim analysis mentions this issue only briefly in its "caveats" to its estimates; the revised analysis discusses catastrophic risk at greater length, suggesting it as an area for future research on the SCC.

In fact, the treatment of catastrophic risk is one of the most important parts of climate economics, and has been the subject of extensive theoretical analysis and debate. Martin Weitzman's important recent work on uncertainty suggests that policy should be directed at reducing the risks of worst-case outcomes, not at balancing the most likely values of costs and benefits.[35] This fits well with a large portion of the prevailing discourse on climate change: The expected damages are important and costly; the credible worst-case outcomes are disastrously greater. The urgent priority is to protect ourselves against those worst cases, not to fine-tune expenditures to the most likely level of damages.

Protection against worst-case scenarios is familiar, though it takes us outside the realm of cost-benefit analysis, into the discussion of insurance policies. Insurance is normally taken out against events which, on average, are unlikely to occur: The average U.S. housing unit can expect to have a fire every 250 years, so the most likely number of residential fires you will experience in your lifetime is zero.[36] Moreover, insurance is guaranteed to fail a simple cost-benefit test – the average value of payments to policyholders must be less than the average value of premiums, in order for any insurance company to remain in business.

Policy designed from this perspective would not be framed in terms of cost-benefit calculations. Rather, it would begin with adoption of a safe minimum standard, based on the scientific analysis of potential risks. The economic analysis would then seek to determine the least-cost strategy for meeting that standard. For example, we recently examined, together with a group of coauthors, the costs of lowering atmospheric CO_2 concentrations to 350 ppm, a level now advocated by a growing number of climate scientists and policy analysts.[37] The best available estimates suggest that the costs would be noticeable, but manageable. The

[34] As cited in EPA (2009), "Proposed Rulemaking."

[35] Weitzman, M. (2009),"On Modeling and Interpreting the Economics of Catastrophic Climate Change," *Review of Economics and Statistics*, 91:1-19; see also Weitzman, M. (2007) "A Review of the Stern Review on the Economics of Climate Change," *Journal of Economic Literature*, 45:703-724. For a non-technical presentation of the Weitzman analysis of uncertainty as applied to climate change, see Ackerman, F. (2009), *Can We Afford the Future? Economics for a Warming World*, London: Zed Books, Chapter 3.

[36] See Ackerman (2009), Chapter 3, for details.

[37] Ackerman et al. (2009), *The Economics of 350* (see full reference above).

risk of spending "too much" on clean energy alternatives pales in comparison with the risk of spending too little and irreversibly destabilizing the earth's climate.

The revised analysis adds complexity, but not insight

The features, and problems, described above are common to both the interim and revised calculations of the SCC. The more elaborate analysis in the revised calculation, used in the small electric motors case, adds two more major features.

First, the working group performed a Monte Carlo analysis of the effects of scientific uncertainty about climate sensitivity. This appears to be done in a rigorous, appropriate manner. One might expect the result to be a much higher SCC, but that is not the case. We made a similar, unexpected discovery in recent research with DICE: Varying the climate sensitivity alone caused surprisingly little change in the model results.[38] In schematic terms, climate sensitivity governs the model's translation of CO_2 concentrations into temperature increases, but the model's damage function translates temperatures into economic damages. If the damage function is sufficiently "optimistic" – or perhaps "Polyanna-ish" – then even relatively high temperatures may impose limited costs, and therefore inspire limited policy responses. The DICE damage function rises at a leisurely pace as the world gets warmer, and does not project that half of global output is lost to climate damages until warming reaches 19°C (34°F), far above the range of temperatures normally considered in even the most disastrous climate scenarios.

Second, the working group chose a set of macroeconomic growth and emissions scenarios for use in the three models. Rather than using relatively familiar IPCC scenarios, the working group opted, with little explanation, for a group of five scenarios extracted from the Energy Modeling Forum 22 (EMF 22) process. EMF 22 compared the latest projections from about a dozen climate economics models (including FUND, but not DICE or PAGE). The working group took four distinct business-as-usual scenarios, from four different EMF 22 models, and one policy scenario achieving moderate emissions reduction. It then used these scenarios in DICE, FUND, and PAGE, and averaged the results. For DICE in particular, significant software modifications were required to reproduce the EMF scenarios.

This hybrid modeling exercise is unsatisfying all around: It has neither the benefits of relative familiarity with the three simple models and the standard IPCC scenarios, nor the advantages of applying the more complex, larger models used in EMF 22. If such large pieces of the EMF 22 apparatus needed to be used, why not review the findings of the EMF 22 models as a whole?

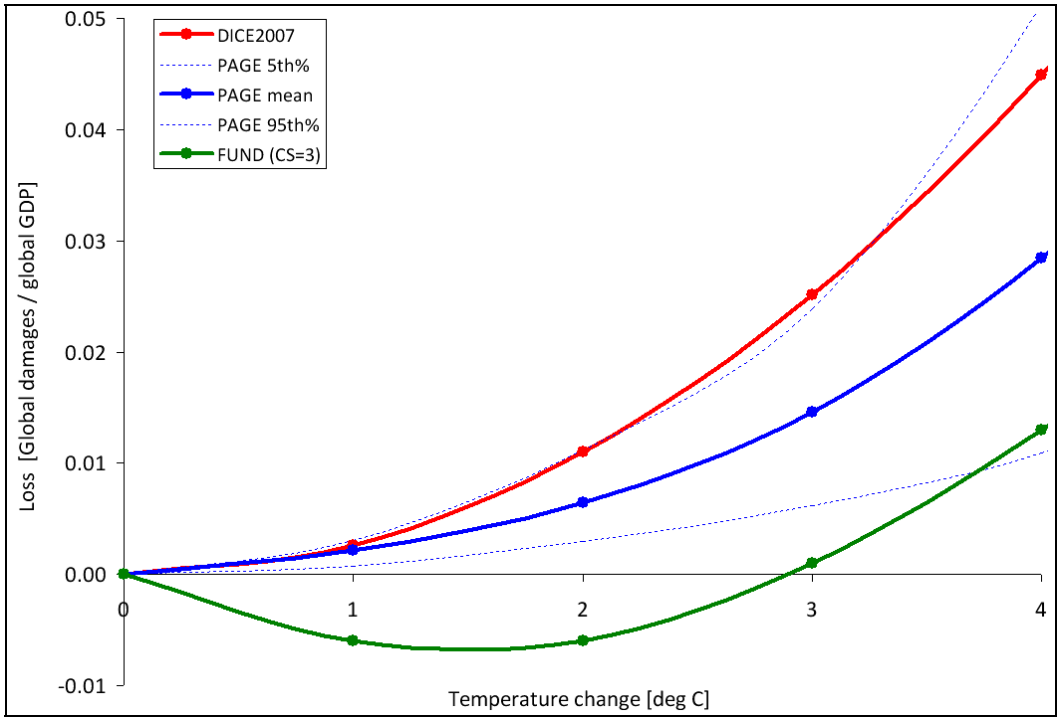
One conclusion from the revised analysis is that FUND is an outlier among climate economics models. As the working group's Figure 1B (reproduced here) shows, DICE and PAGE project modest but positive damages at low temperature changes, but FUND projects net benefits (in the graph above these are represented as negative – below zero – costs) to the world from warming until the world is almost 3°C (more than 5°F) hotter. That is, FUND believes the world will be better off as a result of the first several decades of global warming. With a high enough discount rate, those decades of desirable warmth outweigh the far future when we move beyond 3°C; at a 5 percent discount rate, FUND's estimate of the SCC is negative![39] According to FUND, that is, at a 5 percent discount rate, it would be appropriate to

[38] Ackerman, F., E.A. Stanton, and R. Bueno (2010), "Fat Tails, Exponents, and Extreme Uncertainty: Simulating Catastrophe in DICE," forthcoming in *Ecological Economics.*

[39] See U.S. Department of Energy (2010), Appendix 15A (full reference above), page 27, Table 3; see also EPA (2009) "Proposed Rulemaking," page 49615, Table III.H.6–1, for comparable results in published FUND studies.

provide a (small) subsidy to those who emit carbon, because they are accelerating the arrival of the gloriously hotter mid-century years.[40]

Annual Consumption Loss for Lower Temperature Changes in DICE, FUND, and PAGE



Source: U.S. Department of Energy (2010), Appendix 15A (see full reference above), page 11, Figure 1B.

Our reading of these results is that the FUND model needs to be towed back to the shop for major repairs. The interagency working group, however, has concluded that FUND is an appropriate choice for its short list of three models providing estimates of the SCC for U.S. policy purposes.

The results of the analyses will be interpreted as clustering around $21 per ton

Neither of the analyses resulted in a single bottom-line estimate of the one and only SCC. Both produced multiple figures, primarily reflecting differing assumption about the discount rate. The figures are presented in Table I.

[40] At higher discount rates, FUND's estimates of the SCC move into barely positive territory: $6 per ton at a 3 percent discount rate, and $14 per ton at 2.5 percent, still far below DICE and PAGE. DOE (2010), Table 3, p. 27.

Table I: Estimates of the SCC in two recent analyses

	Social cost of carbon, 2010 (in 2007$)	
	Interim analysis	Revised analysis
Fixed discount rates, mean estimates		
5 percent	$5	$5
average of 3 and 5 percent	$21	–
3 percent	$37	$21
2.5 percent	–	$35
Other estimates		
discount rate declines from 5 percent	$11	–
discount rate declines from 3 percent	$61	–
95th percentile risk, 3 percent discount rate	–	$65

Note: Published SCC values in the interim analysis are for 2007; we have escalated them at the recommended 3 percent annual real growth rate to yield figures for 2010, comparable to the revised values.

Both the interim and revised analysis provided three estimates involving fixed discount rates and mean risks; at first glance the ranges of numbers are very similar. The definitions, however, are different: At a 5 percent discount rate, the SCC is the same for both, but at 3 percent the revised SCC is $21, while the corresponding interim value is $37. At least at lower discount rates, the two analyses embody very different views of the damages caused by greenhouse gas emissions. Due to the accidents of presentation, however, each has a (differently defined) "central" estimate, in this group, of $21.

Each analysis also considered one alternative assumption. The interim analysis examined the implications of discount rates that decline over time, starting at either 5 percent or 3 percent. The revised analysis calculated the 95th percentile risk, from its Monte Carlo analysis of climate sensitivity, using its "central" estimate of a 3 percent discount rate. The uppermost values projected are, coincidentally, not far apart, though again the definitions differ. While the upper values are academically interesting, both analyses are being taken as projecting that the SCC is $21 per ton of CO_2 in 2010, measured in 2007 dollars, or roughly 20 cents per gallon of gasoline. It is hard to reconcile a carbon price that is well within the range of normal price fluctuations at the gas pump with the policy goal of substantially reducing carbon emissions.

5. Conclusion and Recommendations

The administration's narrow proposed range of SCC values, with a likely "central" estimate of $21, is a function of its choice of a limited range of underlying studies, high discount rates, and insufficient emphasis on the risk of catastrophic climate damage. Different choices at several points in the methodology would have resulted in a far higher SCC and, as a result, more stringent and more expensive emissions reduction would be considered economical.

The discussions of the SCC in the working group analyses to date do not contain enough information to construct a better estimate. Instead, there is a need for more extensive research, examining the full range of available studies of climate damages and costs, and analyzing assumptions about the risks and magnitudes of potential climate catastrophes. If one or more of the simple climate economics models highlighted in the rulemaking process –

DICE, FUND, and PAGE – are to be used, then the default data sets supplied by the modelers need to be independently validated against the latest research on climate damages and other input assumptions. FUND, in particular, needs to be re-examined to understand its projections of net benefits from warming, and to consider the potential need for modification.

Additional research on climate damages could address the potential disconnect between science-based and economics-based targets for emission reduction. If, as climate science forecasts with increasing urgency, there are severe risks from just a few degrees of warming, this should be reflected in the economic estimates of damages. Models which claim that the first 3°C of warming are beneficial to the world will inevitably endorse very little reduction in greenhouse gas emissions; models which imply that 2°C is the limit for avoiding serious risks of destabilizing the earth's climate will suggest much greater reduction. They can't both be right. There is no reason to think that three small economic models contain insights into climate dynamics superior to those from the massive, extensively tested general circulation models of climate science. Thus it is time to construct an economic analysis consistent with the concerns and constraints that emerge from the science of climate change.

There is a related need for policies to address the crucial but unmonetized aspects of climate impacts, and to face the ethical choices raised by these impacts. These choices are difficult to fit into the cost-benefit framework implicit in the SCC calculation.[41] An alternative approach could assert that it is essential to set an absolute limit on climate damages, and therefore to keep emissions and temperature increases under strict ceilings – such as 350 ppm of CO_2 in the atmosphere, or no more than a 2°C temperature increase. This would lead to a cost-effectiveness analysis, seeking the least-cost scenario for achieving the needed emission reductions. That scenario would consist of adopting all the lowest-cost reduction opportunities, starting from the measure with the lowest cost per ton of avoided emissions and adopting additional measures in order of their expense. In a cost-effectiveness framework, the carbon price is still important to decision-making, but it is calculated on a different basis. Instead of an SCC, the carbon price would instead represents the per-ton cost of the most expensive mitigation measure that is required to meet the emission reductions target.

How high might an alternative carbon price turn out to be? The United Kingdom, which pioneered the use of SCC estimates for policy purposes, abandoned calculation of the SCC altogether in 2009, and now bases its carbon price on estimates of mitigation costs (as would be required under a cost-effectiveness approach). The latest estimate is a range of $41-$124 per ton of CO_2, with a central case of $83 – which is very close to the estimate of the SCC in the Stern Review.[42] An expanded calculation of carbon prices for the United States should at least explore prices in this range, and should be open to considering the full range of implications of the extensive research that is needed to compute a better estimate of the price of carbon emissions.

[41] Ackerman, F., and L. Heinzerling (2004). *Priceless: On Knowing the Price of Everything and the Value of Nothing*, New York, NY: The New Press.

[42] U.K. Department of Energy & Climate Change (2009), "Carbon Appraisal in UK Policy Appraisal: A Revised Approach," available online at http://www.decc.gov.uk/en/content/cms/what_we_do/lc_uk/valuation/valuation.aspx. Pounds sterling converted to dollars using an exchange rate of £1.00 = US$1.625.

SUGGESTED CITATION:
Frank Ackerman and Elizabeth A. Stanton "The social cost of carbon", *real-world economics review*, issue no. 53, 26 June 2010, pp. 129-143, http://www.paecon.net/PAEReview/issue53/AckermanStanton53.pdf

Keynes, Recovered*

review article by Jonathan Kirshner

The Keynes Solution: The Path to Global Economic Prosperity
Paul Davidson
Palgrave Macmillan, $22 (cloth)

Keynes: The Rise, Fall, and Return of the 20th Century's Most Influential Economist
Peter Clarke
Bloomsbury Press, $2o (cloth)

Keynes: The Return of the Master
Robert Skidelsky
Public Affairs, $25.95 (cloth)

"Keynes is back." It is a familiar cliché, but also an enigma. Enigmatic, first, because Keynes, the most influential economist of the twentieth century, never really left. Like it or not, we live in a macroeconomic world elucidated by Keynes and those who followed in his footsteps. Even Robert Lucas, who won a Nobel Prize for his criticisms of conventional Keynesianism, said in response to the financial crisis: "I guess everyone is a Keynesian in a foxhole."

But enigmatic also because Keynes himself was never with us. From his vast writings, a few ideas were quickly distilled into analytical tools and policy prescriptions that became known as "Keynesianism."

This produced some stark differences between Keynes's ideas and those that bore his name. Once, after a wartime meeting with American economists, Keynes observed, "I was the only non-Keynesian in the room." Following his death in 1946, the divergence only grew. And in due course, even the Keynesian distillate was superseded by shifts in economic theory and overtaken by events, especially the stagflation of the 1970s, the political climate of the 1980s, and the ideological triumphalism of the post–Cold War 1990s. With Keynes himself gone, and Keynesianism losing its grip, the policy choices in the decade leading up to the great financial crisis of 2007-08 were the unhappy but predictable result.

Just what did Keynes believe?

Keynes thought that most markets work well most of the time, and that there is no economically viable substitute (or attractive philosophical alternative) to undirected individuals pursuing their idiosyncratic interests, guided by a well-functioning price mechanism.

But Keynes also knew that some markets don't work some of the time, and a few markets, in particular circumstances, perform dismally. He devoted much of his career to addressing these dysfunctions, with three main concerns: first, that an economy, once stuck in a rut, often will be unable to right itself on its own; second, that financial markets—driven, for better and worse, by herd behavior, uncertainty, and unpredictable shifts in the attitudes and emotions that he called "animal spirits"—are inherently prone to destructive cycles and panics; and third, that an unregulated international monetary system tends to veer toward

* This review article appears here with the kind permission of the Boston Review.

unsustainable disequilibria and always presents inescapable dilemmas of balancing domestic and international monetary stability.

These problems—self-reinforcing economic ruts, financial panics, and global instability—are painfully familiar, and they should remind us why there is much to be gained from rediscovering Keynes: not simply returning to "Keynesianism," but searching out the master himself. The authors of three new books on Keynes are eager, and well-situated, to seize this opportunity.

Paul Davidson is one of the leading scholars of the "post-Keynesian" school. While most American economists moved from Paul Samuelson's "Neo-classical Synthesis" to an ever-more circumspect (modestly interventionist) "New Keynesianism," the post-Keynesians emphasized two of Keynes's ideas that the more conventional view minimized: the importance of uncertainty and the "non-neutrality of money" (the idea that changes in money supply affect growth rates and other real economic variables, and not only prices). Peter Clarke wrote the important study *The Keynesian Revolution in the Making*, which explores Keynes's intellectual journey. In his 1924 *Tract on Monetary Reform*, Keynes shows his gifts, but remains fairly mainstream. By 1930 his increasingly creative prescriptions for the struggling British economy culminate in *A Treatise on Money*. And from there, *The General Theory* (1936), with its comprehensive dissent from the classical school, rooted in Keynes's conclusions about problems inherent in laissez-faire capitalism. Finally, Robert Skidelsky is author of a justly celebrated three-volume biography of Keynes.

Each scholar's new book reflects his distinct intellectual engagement with Keynes, but collectively they point in a common direction. Keynes, they suggest, would argue that our current mess has three main causes: failures of the economics profession, mistakes by government, and regrettable social trends.

To appreciate the charge about the economics profession, we need to understand the practical and analytical vulnerabilities of 1960s Keynesianism. On the practical side, Keynesianism foundered on the supply shocks of the 1970s. On the analytical side, Milton Friedman and Edmund Phelps pointed out that Keynesian models failed to take into account that individuals adapt their expectations to new circumstances, including policy interventions, and that those adaptations might prevent Keynesian policies from achieving their desired results.

In the 1980s and '90s New Keynesianism sought to respond to this criticism by integrating "rational expectations" into its analysis. But as the economist James Tobin complained, this was New Keynesianism without Keynes. In particular, Keynes did *not* assume any kind of rational expectations—the idea that individuals, armed with a (shared) knowledge of the (singularly correct) model of how the economy functioned, would instantly, efficiently, and dispassionately optimize choices in an information-rich environment. On the contrary, he emphasized the central role of animal spirits, of daring and ambitious entrepreneurs taking risks and placing bets in a highly uncertain economic environment characterized by crucial unknowns and unknowables.

Uncertainty, as Clarke argues in *Keynes: The Rise, Fall, and Return of the 20th Century's Most Influential Economist*, is a "guiding insight at the heart of Keynes' intellectual revolution." Davidson and Skidelsky also drive the point home in *The Keynes Solution: The Path to Global Economic Prosperity* and *Keynes: The Return of the Master*, respectively. Keynes identified his treatment of expectations about the future as a principal point of departure from the reigning perspective. "The orthodox theory assumes that we have a

knowledge of the future of a kind quite different from that which we actually possess," Keynes explained. "This hypothesis of a calculable future leads to a wrong interpretation of the principles of behavior . . . and to an underestimation of the concealed factors of utter doubt, precariousness, hope and fear."

Many mainstream economists came to embrace the idea of rational expectations and its partner the "efficient markets hypothesis," which holds that current market prices accurately express the underlying value of an asset because they reflect the sum of a collective, rational calculus. Keynes, however, held that investors more often grope in the dark than calculate risk—they can't assign precise probabilities to all potential eventualities because too many factors are unknowable. Faced with this uncertainty, investors inevitably place great weight on the apparent expectations of others. Thus, as Keynes famously argued, investors need to make their best guesses not simply about the likely business environment, but also about the guesses of other investors regarding that environment. The resulting herd-like behavior can at times generate dysfunctional consequences, such as self-fulfilling prognostications of financial crisis.

Embracing rational expectations and the efficient markets hypothesis led the economics profession to make disastrously misguided assumptions about the stability and self-correcting hyper-rationality of financial markets in particular. The government's failure was to embrace these errors of the economics community and run with them. Davidson holds that the origin of the current crisis "lies in the operation of free (unregulated) financial markets." "Liberalized financial markets," he argues, "could not heal the bloodletting that they had caused."

Those markets did not liberalize themselves, of course. Since the 1980s public policy has embraced with increasing momentum the idea that once set free, financial markets take care of themselves. Thus the push by the International Monetary Fund (IMF) in the mid-1990s to force states to eliminate their capital controls, and the 1999 repeal of the Glass-Steagall Act: both instances of dismantling regulations designed to avoid the horrors of the Great Depression. Glass-Steagall built firewalls to prevent another collapse of the banking system; prudent capital controls aimed to mitigate the tendency of the international monetary system to transmit destabilizing shocks across borders. "Nothing is more certain than that the movement of capital funds must be regulated," Keynes insisted, as one of the key architects of the IMF.

These books suggest that Keynes also would have found something troubling about contemporary Western society, or at least the version of it that emerged in the post–Cold War United States. Keynes's economics was rooted, always, in a normative, philosophical vision, in particular that the economic game was not about winning the candle. This was one of Skidelsky's great contributions in his biography, and *Return of the Master* treats us to a chapter on "Keynes and the Ethics of Capitalism," including Keynes's aversion to "the objectless pursuit of wealth." Rather, the purpose of economic activity (and the role of the economist) was to "solve" the economic problem—the provision of adequate needs, opportunities, and comfort—so that the important things in life could be pursued.

But in the mania of the American housing bubble, the chase of wealth became everything. The financial sector expanded, national savings rates plunged, and Clinton-era deregulations were followed by the Bush administration's abdication of government oversight. Financiers eagerly accepted the open invitation to recklessness and enjoyed astronomical compensation packages for embracing imprudent risks. Borrowers took on debt far beyond any responsible expectation of what they could repay. In retrospect even Alan Greenspan finally understood the errors of the era of unregulated finance he championed.

Davidson's *The Keynes Solution* is half of a great book. Its first three chapters tell the "how we got here" story. They summarize clearly and convincingly the pathologies of unregulated finance, and in particular how deregulation of the banking sector skewed incentives and encouraged the chase for speculative profits via increasingly complex and innovative financial assets. Rather than concentrate on the more mundane business of traditional banking, "investors spent vast sums buying . . . mortgage-backed financial securities despite the fact that no one was sure of what actual real assets were pledged as collateral against them." These investors don't sound like the agents modeled on the chalkboards of the "New Classical" macroeconomists, and Davidson points out that Keynes's theories about expectations and uncertainty capture human activity far more accurately than do hypotheses of rational expectations and efficient markets.

After this sparkling display of diagnostic analysis come two chapters of Keynesian responses to the crisis. In pitching fiscal stimulus and monetary expansion, Davidson is pushing a bit on an open door, even as he neatly links these measures back to Keynes's analyses. But Davidson is perhaps too cavalier about the potential long-term costs of stimulus, and he dismisses the risk of inflation via an unconvincing appeal to the non-neutrality of money. It is easy to agree with his fiscal and monetary responses to the crisis: the house was on fire, and that fire had to be put out. But that does not mean there won't be water damage to be reckoned with, and in the looming "out years," when the economy is in recovery, inflationary pressures and burdensome debts will demand close attention.

What about the decades ahead, once recovery is achieved? Davidson advances proposals for the re-regulation and supervision of the financial sector, which follow directly and reasonably from the first part of the book. But his support for protectionist trade policies is irrelevant as a response to the current crisis and based on a very selective and distorted reading of Keynes's positions. Davidson does a disservice to his important early arguments by spending down his own credibility.

A chapter on reforming the world's money is truer to Keynes's proposals for an international clearing union, and international macroeconomic imbalances were certainly an expression of problems related to the crisis. (Keynes anticipated these dilemmas and hoped a more capacious IMF might mitigate them by placing greater burdens of adjustment on creditors.) But one would have expected much more than two pages of afterthought devoted to creative, market-friendly schemes to slow the flow of speculative capital, such as a (nice, Keynesian) Tobin tax, which would inhibit counterproductive currency trades. In sum, despite the strong—if occasionally score-settling—appendix on the disappointing postwar trip from Keynes to Keynesianism, Davidson's book limps to the finish line. A pity, as its first five chapters should be required reading.

Clarke's *Rise and Fall* is also a book in two halves, but it sails upon smoother (and somewhat safer) narrative seas. The first half is a very readable biographical essay on this extraordinarily rich and full life, with an emphasis, perhaps at times excessive, on debunking various Keynesian myths inappropriately associated in the popular imagination with Keynes himself (for example, that Keynes did not mind inflation, or was always on the side of deficits). The second half reprises Clarke's previous tracing of Keynes' intellectual journey of the '20s and '30s, from a gifted and innovative but still classically trained economist into, well, Keynes: the iconic historical figure, the finished product who, having finally escaped from his classical upbringing, told fellow economist Dennis Robertson, "I am shaking myself like a dog on dry land."

Clarke bemoans the divergence between Keynes and the Keynesians, and he tangles with Roy Harrod, Keynes's first biographer. According to Clarke, Harrod's *The Life of John Maynard Keynes* "covered up" the depth of Keynes's opposition to conscription during the first world war and was too circumspect about aspects of his personal life (both true, but most modern Keynes biographers feel ritually compelled to critique this valuable book).

The score settling continues, as Clarke reminds (or informs) the reader that Keynes's comment "in the long run we are all dead" has been egregiously misinterpreted. He was not devaluing the future, but rather calling attention to the fact that it could take a *very long time* for markets to restore equilibrium, and there would be much loss and suffering in the waiting. Similarly, Keynes was not "inconsistent," but was admirably flexible, possessing a quality now in painfully short supply. He was open to changing his mind when the evidence suggested that he should, and to changing his policy prescriptions when circumstances demanded. Clarke also reviews Keynes's allergy to Marxism and his "lifelong belief in the unique virtues of the market," clearing away much of the nonsense that has been said about Keynes over the years.

In detailing the emergence of the Keynesian revolution, Clarke mines Keynes's performance as a member of the Macmillan Committee (the formal British Government inquiry into the causes of the Great Depression), his autumn 1932 professorial lectures on "The Monetary Theory of Production," and his vital intellectual collaborations in the 1920s and 1930s. *The General Theory* was published in late 1936, but by autumn of 1932, Clarke argues, "Keynes had seized upon the essentials of the theory of effective demand" —one of Keynes's critical contributions, which holds, among other things, that downwardly spiraling consumer demand is not self-correcting and can therefore induce long economic dry spells. Reviewing the intellectual, economic, and political debates of the times, Clarke puts together the clues—uncertainty, the precariousness of knowledge, the crucial role of market sentiment—and then unmasks the killer: "Here was Keynes' revolution," he exclaims, "the economy was at rest. It was not self righting."

Davidson and Clarke maintain that Keynes's insights are necessary for understanding today's crisis, overcoming it, and preventing its recurrence. In Skidelsky's *Return of the Master*, this theme is even more explicit, and the arrows from Keynes's theories to today's policies are brightly drawn. This excellent, much-needed book begins with an overview and interpretation of the current crisis; then introduces Keynes, his life, and his life lessons; and concludes with what it would mean, as a practical matter, to apply Keynes's theories.

Like Clarke, Skidelsky is keen to puncture anti-Keynesian myths ("it may surprise readers to learn that Keynes thought that government budgets should normally be in surplus"); with Davidson, he emphasizes disequilibria in the international financial economy. He endorses Davidson's updated Keynesian scheme for an international clearing union. And with the others, Skidelsky emphasizes the central role of uncertainty in Keynes's economics, but he gives particular attention to the philosophical foundation of the framework: the purpose of economics is to allow people to live "wisely, agreeably, and well."

What went wrong? For Skidelsky, the deregulations, especially the elimination of Glass-Steagall, allowed commercial banks to become "highly leveraged speculators in the newly developed securities, with a hubris given by their faith in their 'risk-management' models." Initially, under the guise of not playing the blame game, Skidelsky hands out a

number of get-out-of-jail-free cards, absolving, at least partially, bankers, credit rating agencies, central bankers, regulators, and governments.

This magnanimity allows Skidelsky to focus on his preferred culprit: "the intellectual failure of the economics profession." And his critique is well done. He skewers the New Keynesians for giving up the store to the New Classicals, and both for their embrace of rational expectations and the efficient-markets hypothesis.

But even if one agrees with Skidelsky, his murderer's row of unindicted co-conspirators deserves more blame than it is assigned. Greenspan, for example, can be cited as a living embodiment of the anti-Keynesianism that caused the meltdown. In response to the Asian Financial Crisis of 1997-1998—the previous crisis of unregulated capital—Greenspan's conclusion was that "one consequence of this Asian crisis is an increasing awareness in the region that market capitalism, as practiced in the West, especially in the United States, is the superior model."

This was wrong, but it did reflect Greenspan's cheerleading for a deregulated financial sector that he felt no need to chaperone. Greenspan's support also lent legitimacy to the Bush Administration's ill-advised tax cuts, which made today's problems worse. As Keynes wrote, "the boom, not the slump, is the right time for austerity at the Treasury." The budget surpluses that Bush frittered away should have been used to pay down the federal debt. Instead the debt mushroomed, making the government less able to issue the debt that is now needed.

Skidelsky does eventually come around to a bit of finger-pointing. In addition to the mistakes of economists, he identifies three broader failures: institutional, intellectual, and moral. Banks "mutated from utilities into casinos"; regulators and policymakers "succumbed to . . . the view that financial markets could not consistently misprice assets and therefore needed little regulation"; and society as a whole embraced "a thin and degraded notion of economic welfare." Moreover, although the right emergency measures were taken to stop the economic bleeding, Skidelsky notes that little, so far, has been done to address the underlying problems. Unless we fix them, the "world is likely to limp out of this crisis to the next."

All this sounds right. But what are the solutions? For Keynes, via Skidelsky, they start with taming (that is, re-regulating) finance, and, ambitiously, reforming the way economics is taught: decoupling the teaching of macro- from microeconomics and broadening training so that economists are once again, as Keynes described, parts "mathematician, historian, statesman and philosopher." Is this going to happen? Skidelsky is right to doubt it, since both of his reforms take on powerful, entrenched interests and deeply rooted institutional traditions. "But it may come about gradually," he concludes, "with the shift in power and ideas" away from the United States.

Arguably the crisis has accelerated that shift, and certainly actors in East Asia, now twice burned by unregulated finance, might be quicker to embrace Keynes. This may not offer much hope for American public policy, but it is a provocative thought.

SUGGESTED CITATION:
Jonathan Kirshner, "Keynes, Recovered", *real-world economics review*, issue no. 53, 26 June 2010, pp. 144-149,
http://www.paecon.net/PAEReview/issue53/Kirshner53.pdf

Two comments on

Marvin Brown, "Free enterprise and the economics of slavery"

Did Smithian economics promote slavery?

Bruce Elmslie [University of New Hampshire, USA]

In a recent article published in this *Review*, Marvin Brown claims that Adam Smith's account of wealth creation failed to acknowledge slavery as a major source of the wealth of nations. The result of this was "a dissociative economics that splits off the misery of the actual providers of wealth from the experiences of enjoying it. This is the legacy of the Scottish Enlightenment, and Smith is its best illustration." (Brown, 2010:38) In building to this conclusion, Brown makes three claims. First, from his experience living among Glasgow tobacco traders profiting from products made by slaves, Adam Smith knew of the relationship between slavery and wealth creation but failed to incorporate this knowledge into the *Wealth of Nations*. Second, "[f]rom an economic point of view, slaves are property" (ibid: 35) and "[f]or Smith, the economics of property always overrides the rights of humans…" (ibid: 37) Third is a point of guilt by association where Brown strongly implies that Smith believed that slaves were inferior. This is done within the general context of understanding a specific quote from the *Wealth* where he states that Smith may have failed to discuss the link between slavery and wealth creation "because he could not totally separate the political or moral dimension from the economic." (ibid: 37) This quote will be examined in detail below.

In this comment, I will address each of these points in order. The purpose of this exercise is to give an alternative answer to Brown's main question regarding Smith's lack of recognition of slavery as a major source of wealth creation and its dissociative effects.

Slavery and Wealth Creation

Brown takes it as an obvious fact that "Africans were the providers for much of the wealth for the Atlantic trading nations." (ibid: 38) While one can easily question the actual importance of the slave trade in the wealth of these nations (see Acemoglu, et al., 2005), the important question for our purpose is to ask if Smith believed this to be true. Based on Smith's actual text, the answer seems clearly to be no.

Smith believed that production based on slavery was less productive and more costly than production based on free labor. Smith made three arguments in support of this conclusion: slaves do not produce more output than was necessary to provide for their own subsistence, slaves are not inventive, and freemen are responsible for their own "wear and tear" while slaves are not. Interestingly, and related to Brown's third point, Smith actually insisted that neither of the first two conclusions was based on the natural inferiority of slaves. *Para contra*, they were based on slaves not receiving proper incentives to be productive or inventive. Smith argued that slaves don't work as hard as free men because they have no ability to own property in order to improve their condition. For Smith, this desire was taken to be part of human nature, an axiom not needing explanation. "[T]he principle which prompts to save, is the desire of bettering our condition, a desire which, though generally calm and dispassionate, comes with us from the womb, and never leaves us till we go into the grave." (Smith, 1981: 341) While slaves have this innate desire, they are specifically precluded from saving. This difference in incentives creates a less productive worker. Smith's specific reference was to farming. "[T]enants, being freemen, are capable of acquiring property, and having a certain proportion of the produce of the land, they have a plain interest that the

whole produce should be as great as possible, in order that their own proportion may be so. A slave, on the contrary, who can acquire nothing but his maintenance, consults his own ease by making the land produce as little as possible over and above that maintenance." (Smith, 1981: 389)[1] As Brown points out, slavery in Europe had largely ended in the sixteenth century. Smith who specifically referred to this as "so important a revolution" (ibid) actually used this difference in productivity as one of two major contributing factors in ending slavery, it was in the economic interests of the "proprietor" to see it eliminated.

For Smith there were two major factors in determining wealth creation, the productivity of labor and the proportion of productive to unproductive labor. In focusing on the former, Smith famously cited extensions in the division of labor as determining labor productivity, and much of this was the result of innovations made by the workers themselves. According to Smith, slaves were less inventive. "Slaves, however, are very seldom inventive; and all the most important improvements, either in machinery, or in the arrangement and distribution of work which facilitate and abridge labour, have been the discovery of freemen." (Smith, 1981: 684) Why are slaves not prone to inventiveness? Are they, somehow less capable of discovery than other workers? Smith's answer was a clear no. The incentives of the slave were in the opposite direction. The slave was likely to be punished for showing any ingenuity at all.

> Should a slave propose any improvement of this kind, his master would be very apt to consider the proposal as the suggestion of laziness, and a desire to save his own labour at the master's expence. The poor slave, instead of reward, would probably meet with much abuse, perhaps with some punishment. In the manufactures carried on by slaves, therefore, more labour must generally have been employed to execute the same quantity of work, than in those carried on by freemen. The work of the former must, upon that account, generally have been dearer than that of the latter. (Smith, 1981: 684)

Finally, Smith argues that slaves are more costly because the person in charge of ensuring the health and wellbeing of the slave is not the slave himself. Under slavery, the "wear and tear" of the slave is "commonly managed by a negligent master or careless overseer". The freeman, on the other hand managed his own wear and tear. This difference results in increased expense in keeping a slave healthy enough for productive work. As a result, the labor of freemen was less costly than that of slaves even when wages are high. "It appears, accordingly, from the experience of all ages and nations, I believe, that the work done by freemen comes cheaper in the end than that performed by slaves. It is found to do so even at Boston, New York, and Philadelphia, where the wages of common labour are so very high." (Smith, 1981: 99)

Thus, a simple answer to our main question is forthcoming after addressing this first point only. Smith did not discuss slavery as an important element in the growth of wealth because, he regarded the institution as detrimental to wealth accumulation. Simply put, slave labor was considered to be more costly than that of free labor. In each case he addressed, slavery as an institution produced an incentive structure that decreased worker productivity, inventiveness and health. As Smith states, "the work done by slaves, though it appears to

[1] In a similar passage Smith again states the connection between the inability of slaves to acquire property and the minimization of work effort, "A person who can acquire no property, can have no other interest but to eat as much, and to labour as little as possible. Whatever work he does beyond what is sufficient to purchase his own maintenance, can be squeezed out of him by violence only, and not by any interest of his own." (Smith, 1981: 387-88) Note that in both quotes, slaves are making a perfectly rational assessment of their own interests.

cost only their maintenance, *is in the end the dearest of any*." (Smith, 1981: 387, emphasis added)[2]

Private Property Trumps Human Rights

Brown argues that the economics of Smith held a strong implicit argument in support of slavery. In support of this argument, Brown takes a quotation from Smith's *Lectures on Jurisprudence* where he makes the case that as societies develop and become more complex "the greater will be the number of their laws and regulations necessary to maintain justice, and prevent infringement of the right to property." (LJ, Brown, 2010: 37) Brown utilizes this passage to make the following claim: "Smith never mentions the role of slavery in the commercial society he enjoyed, but here we do see how important it was that there were laws to protect an owner's property, or in the case of slavery, to protect the slave owner." (Brown, 2010: 37) The question to be addressed here is did Smith believe that the government should protect the rights of slave owners over slaves?

To address this question, we should begin with a more general relationship between property rights and freedom and then move to the specific question of the sovereign's role in supporting the property rights of slave owners. It is clear from many passages in the *Wealth of Nations* that Smith was a strong believer in private property as being fundamental in the accumulation of national wealth. However, the right to own property was never universal. It had limits. One such limit was infringing on the rights of others. In this sense it is not the case that "the economics of property rights always overrides the rights of humans." Smith makes this case most clearly when addressing the heart of his system of natural liberty. "Every man, *as long as he does not violate the laws of justice*, is left perfectly free to pursue his own interests his own way…" (Smith, 1981: 687, emphasis added)

Thus, the question becomes, do the laws of justices apply to slaves? To get an answer, we can look to see how Smith viewed the decline of slavery in Europe, and the proper role of government in ending it. In fact, Smith was supportive of the ban and endorsed the strong role that he saw governments (Monarchs) playing in ending slavery. He actually took the opportunity to criticize the Catholic Church for both claiming too great a role and condemning slavery too little.

> The time and manner, however, in which so important a revolution was brought about, is one of the most obscure points in modern history. The church of Rome claims great merit in it; and it is certain that so early as the twelfth century, Alexander III. published a bull for the general emancipation of slaves. It seems, however, to have been rather a pious exhortation, than a law to which exact obedience was required from the faithful. Slavery continued to take place almost universally for several centuries afterwards, till it was gradually abolished by the joint operation…of the proprietor on the one hand, and that of the sovereign on the other. (Smith, 1981: 389-90)

Smith is clearly arguing that a proper role of the sovereign, in the general administration of laws of justice, was the establishment of laws outlawing slavery. In no way did Smith or his system of natural liberty—including the right to own property—endorse the institution of slavery. This point is made more clearly in his *Lectures on Jurisprudence* where Smith argues that the interests of slave owners make it more difficult for slavery to be ended in democracies. In such a case, Smith argued that "the rights of humans" trumps both

[2] In the *Lectures on Jurisprudence* Smith made similar arguments. One in particular refers to the American colonies directly: "Our colonies would be better cultivated by freemen." (Smith, 1982: 453)

freedom by which he means democratic freedoms of land owning voters and wealth. "Freedom and opulence contribute to the misery of the slaves. The perfection of freedom is their greatest bondage. And as they are the most numerous part of mankind, *no human person will wish for liberty in a country where this institution is established*." (Smith, 1982: 453, emphasis added)[3] For Smith basic humanity trumps the liberty of property owners when it comes to slavery.

The Inferiority of Non-Europeans

In raising the question of the inferiority of slaves, Brown quotes Smith as follows:

> The pride of man makes him love to domineer, and nothing mortifies him so much as to be obliged to condescend to persuade his inferiors. Wherever the law allows it, and the nature of the work can afford it, therefore, he will generally prefer the service of slaves to that of freemen. (Smith, 1981: 388)

Directly following this quote, Brown asks:

> How are we to understand this explanation? When Smith speaks of 'the pride of man,' does he have the tobacco lords in mind? Are these 'men' members of the political economic club he attended in Glasgow? We don't know. We do know that Smith lived in a world where it was common to see American, Africans, and Asian as inferior to Europeans. Still, the terminology of superior and inferior places both groups in the same species, instead of different types of things: humans and property. Perhaps the key here is the law. (Brown, 2010: 38)

It turns out that we actually do have a good idea who Smith meant both in his reference to the pride of man and in his use of language. He was not speaking of tobacco traders, but slave owners, and his use of the word inferior was a reference to the perception of slaves being inferior by these very slave owners. This additional clarity comes from his *Lectures on Jurisprudence* where he covered similar territory using similar language. Moreover, the law which Smith was referring to in the quotation above was the object of his criticism. The economics of slavery did not pay in the form of higher profits; it was the pride of man aided by the law that helped to maintain slavery at the expense of additional profit.

> In a democraticall government it is hardly possible that it [slavery] ever should, as the legislators are here persons who are each masters of slaves; they therefore will never incline to part with so valuable a part of their property; and tho as I have here shewn *their real interest would lead them to set free their slaves and cultivate their lands by free servants* or tenents, yet the love of domination and authority and the pleasure men take in having every <thing> done by their express orders, rather than to condescend to bargain and treat with *those whom they look upon as their inferiors* and are inclined to use in a haughty way; *this love of domination and tyrannizing, I say, will make it impossible for the slaves in a free country ever to recover their liberty*. (Smith, 1982: 186, emphasis added)

In fact, Smith gives a similar account in the extended paragraph containing the "pride of man" sentence quoted above. Slavery existed as the dominant mode of production in tobacco and sugar because they were particularly profitable, not the other way around. In

[3] The same material from the student report of 1762-63 is depicted as follows: "A humane man would wish therefore if slavery has to be generally established that these great blessings [opulence and freedom], being incompatible with the happiness of the greatest part of mankind, were never to take place." (Smith, 1982: 185)

fact, the proportion of slaves to freemen in production was directly related to the profits of the industry because high profits allowed these human proclivities to be satisfied. "Both [sugar and tobacco] can afford the expence of slave-cultivation, but sugar can afford it still better than tobacco. The number of negroes accordingly is much greater, in proportion to that of whites, in our sugar than in our tobacco colonies." (Smith, 1981: 389)

This statement leaves open to question where, if not from the use of slaves, do these higher profits originate? In a more general discussion, not directly tied to the slavery question, Smith tells us that these profits are due to monopoly trading rights (Smith, 1981: 174-5) which he argues forcefully against in Book IV of the *Wealth*. However, in an early draft of the *Wealth of Nations*, Smith makes the connection more directly in the context of an argument condemning these profits. In fact, Smith makes it abundantly clear that the proprietors and their communities pay for slavery in the form of decreased wealth even though monopoly trading rights make the cultivation of sugar and tobacco more profitable than commodities that do not enjoy such protection.

[Cultivators of sugar and tobacco] who thus enjoying a sort of monopoly against all the rest of the world, *indemnify* themselves by the exorbitancy of their profites for their expensive and thriftless method of cultivation. The great expence of slave cultivation in the sugar plantations. The yet more exorbitant profites of the planter. That the planters in the more northern colonies, cultivating chiefly wheat and Indian corn, by which they can expect no such exorbinant returns, find it not for their interest to employ many slaves, and yet *Pensilvania, the Jerseys, and some of the provinces of New England are much richer and more populous than Virginia, notwithstanding that tobacco is by its ordinary high price a more profitable cultivation.* (Smith, 1982: 579-80, emphasis added)

In the end, there is no evidence, direct or indirect, that Smith believed that slaves or non-Europeans generally were inferior to Europeans. As we have seen from the argument made previously, Smith did believe that slaves were less productive and inventive compared with freemen. However, the reasoning was always based on incentives rather than natural abilities. Did Smith infer that slave owners were superior to slaves? Again the evidence is no. Slave owners may view slaves as "their inferiors" but these owners must pay dearly for this "love of domination" in the form of lower monetary profits. Moreover, this payment extends to non-slave owners as well. In areas where slavery has been abolished, the standard of living was higher because slavery had been ended.

Conclusion

Near the end of his article, Brown states that the "blind optimism of Smithian economics depends on ignoring the desperation and powerlessness of those who are used to produce goods and services…" (Brown, 2010: 38) When it came to slavery, Smith was neither blind nor particularly optimistic. For Smith, "[t]he freedom of the free was the cause of the great oppression of the slaves." (Smith, 1982: 182) And this reality was naturally reinforced by the very self interests that, in another context, resulted in the social harmony epitomized so famously by his imagery of the invisible hand. When beginning with a basic injustice, the interests of the oppressor is to create laws and norms that maximize oppression. As stated previously, democracies were considered the strongest vehicle for driving this vicious circle. In such a system "whatever laws are made with regard to slaves are intended to strengthen the authority of the masters and reduce the slaves to a more absolute subjection. The profit of the masters was increased when they got greater power over their slaves." (Smith, 1982: 181) This is Smithian economics. Self interest promotes general wellbeing when constrained by justice and competition. When beginning with an injustice

such as slavery or a lack of competition from the monopoly power of a capitalist, self interest plays a strong detrimental role in the determination of social wellbeing. Simply put, origins and social context mattered for Smith. He did not devise the world of Pareto efficiency where we take the initial distribution of wealth and power as given. In many instances, how to overcome this initial social structure was the economic question to be answered.

References

Acemoglu, Daron, Simon Johnson, and James Robinson "The Rise of Europe: Atlantic Trade, Institutional Change, and Economic Growth," *American Economic Review*, Vol. 95, No. 3, 2005, pp. 546-579.

Brown, Marvin "Free Enterprise and the Economics of Slavery," *Real-World Economics Review*, Issue No. 52, 7, March 2010, pp. 28-39.

Smith, Adam *Lectures on Jurisprudence*, ed. R. L. Meek, D. D. Raphael, and P. G. Stein, Indianapolis: Liberty Classics, 1982.

Smith, Adam *An Inquiry into the Nature and Causes of the Wealth of Nations*, ed. R. H. Campbell, and A. S. Skinner, Indianapolis: Liberty Classics, 1981.

Author contact: bte@unh.edu

SUGGESTED CITATION:
Bruce Elmslie, "Did Smithian Economics Promote Slavery?", *real-world economics review*, issue no. 53, 26 June 2010, pp.150-155, http://www.paecon.net/PAEReview/issue53/Elmslie53.pdf

Adam Smith's real views on slavery: a reply to Marvin Brown

Thomas Wells [Erasmus University, Netherlands]

Marvin Brown's recent paper in the *Real-World Economics Review* (52) misuses Adam Smith gravely in making his argument that Smithian economics was responsible for slavery. I consider this particularly unfortunate since Adam Smith was certainly a real-world economist and thus a natural friend to the supporters of this journal.

In this short response I will confine myself to rebutting Brown's uses and interpretations of Smith rather than addressing the wider argument of the paper. I would note however that it is not uncontroversial whether slavery was economically more productive than employing 'freemen' (Smith certainly thought not), even though there was certainly profit to be made in the slavery business. This reflects a difference between business and economics – that some people may have made lots of money selling cigarettes says little about its general economic benefits to society.

Brown's argument proceeds by contrasting the facts of 18th century commercial society with Adam Smith's economic theory to reveal that Smith committed sins of omission (neglecting to mention the role of slavery in the British economy) and commission (his economic theory endorsed slavery by prioritising property rights above all).

Smith's sins of omission

> As we know, Smith never mentions the role of slavery in the commercial society he enjoyed... (Brown 37)

> It is hard to know how much [Smith] knew about the plight of slaves on the tobacco or sugar plantations, or how much his readers wanted to know. (Brown 36)

On closer and perhaps more neutral inspection of Smith's work Brown's claims seem hard to sustain. Smith actually talks about slavery quite a lot in the *Wealth of Nations* [WN] (and extensively in his L*ectures on Jurisprudence* [LJ], though he never published these himself). His remarks can be divided into an analysis of the political-economy, ethics, and empirics of slavery.

Smith argues that slave labour is generally less productive than free labour since freemen work for themselves and will therefore work much harder and better for less, so it is not generally in the interest of slave-owners (for example, WN I.8.40). He also notes that slavery reduces productivity growth because slaves do not have any incentive to innovate and argues that this fact has held back societies' economic development throughout history (WN IV.9.47). (However he does acknowledge that the high profit tobacco and even higher profit sugar plantations were able to afford the additional expense of slave labour, particularly if there was a shortage of free labour available (WN III.2.10).) These are hard-nosed arguments meant to persuade hard-nosed readers – they do not concern justice but they do forestall arguments for slavery based on utilitarian considerations.

As to the ethics of slavery, Smith is clearly against it. Brown himself notes this (Brown 33) but allows the reader to fall under the misleading impression that in this Smith's feelings went no further than was fashionable, and anyway found no place in his "dissociative",

property rights based economics.[171] However, what one in fact sees is that Smith assaults the foundations of slavery on multiple fronts.[172] His distaste and disapproval of slavery permeates his economic analysis. For example, "In every country where the *unfortunate law* of slavery is established.... (WN IV.7.76)." It is true that most of Smith's moral analysis of slavery comes in his analysis of ethics *TheTheory of Moral Sentiments* (TMS) and government (LJ), but since most scholars consider that Smith was engaged in a coherent project to systematise different branches of philosophy, that would seem to fall under division of labour rather than schizophrenia.[173]

Thus we find that Smith wasn't content to express a merely passive distaste for slavery. Smith paints vivid pictures of the barbarity of slavery which invite our sympathy with the slave and the justified resentment he should feel: the severity and arbitrariness of punishment (LJ 181); lack of family rights (LJ 178), fatal neglect of their children (LJ 193).

Smith staunchly criticises the ethical failings of those who own slaves: only the kind of 'proud' people who like to dominate others (without having to persuade them with words, or money). "The pride of man makes him love to domineer, and nothing mortifies him so much as to be obliged to condescend to persuade his inferiors. Wherever the law allows it, and the nature of the work can afford it, therefore, he will generally prefer the service of slaves to that of freemen (WN III.2.10)". Brown professes to be puzzled by what Smith can mean by this, but taken in the context of Smith's concerns and methodology it seems to be the answer to the puzzle of why the economically inefficient institution of slavery is so pervasive in human history and the world Smith lived in – it springs from a natural tendency in human nature (pride) that easily becomes a vice (domination) without intervention. That tendency is, as Smith makes clear, particularly apparent in the richest and most powerful merchants and aristocrats who are in fact most secure from the humbling domain of market competition where their partiality in favour of themselves would be ridiculed (TMS 1.iii.3.5).

Smith also savagely criticises the character of those who trade in slaves:

There is not a negro from the coast of Africa who does not.... possess a degree of magnanimity which the soul of his sordid master is too often scarce capable of conceiving. Fortune never exerted more cruelly her empire over mankind, than when she subjected those nations of heroes to the refuse of the jails of Europe, to wretches who possess the virtues neither of the countries which they come from, nor of those which they go to, and whose levity, brutality, and baseness, so justly expose them to the contempt of the vanquished. (TMS V.I.19)

At various points Brown suggests that Smith knew little about the significance of slavery for his commercial society, or perhaps tried to hide that knowledge. But WN and especially Smith's Lectures are filled with detailed notes on slavery in history and his contemporary world. He notes for example recent developments in New England (manumission by the Pennsylvania Quakers (WN III.2.10); and the latest legal interpretations on the status of slaves brought to Britain (LJ 456).[174]

---

[171] Brown's phrasing "It is well known that Adam Smith was against slavery" (Brown 33) is inapt. In fact Smith was famous for his anti-slavery positions (Griswold 1999, 198)

[172] For an excellent account of how Smith goes about this see Griswold pp 198-201

[173] See for example Charles Griswold's admittedly bold reconstruction of Smith's intended corpus, including missing and uncompleted parts (Griswold 1999, chap. 1).

[174] Brown remarks that Smith must have known about such cases but doesn't mention them in WN, implying that he should have. This kind of criticism doesn't seem fair to Smith who after all was writing a

In judging Smith's coverage of the Africa-Americas trade one should also bear in mind the scale of slavery. In Smith's contemporary world "A small part of the West of Europe is the only portion of the globe that is free from slavery.....and is nothing in comparison with the vast continents where it still prevails (LJ 96)," while history, even of the 'civilised' Greek and Roman states, reinforced the apparent pervasiveness of slavery whether within or without the commercial society Smith lived in.

Smith's sins of commission

[T]he four stages are as much a story of property and property relations as a story of the evolution of the means of production..... here we do see how important it was that there were laws to protect an owner's property, or in the case of slavery, to protect the slave owner. For Smith, the economics of property always overrides the rights of humans, and especially the rights of those who did not belong to "commercial society.' (Brown 37)

The blind optimism of Smithian economics depends on ignoring the desperation and powerlessness of those who are used to produce goods and services, whether they are slaves, workers, women, children in sweatshops, or illegal immigrants. (Brown 38)

One of the problems of reading Smith through a particular lens – such as slavery – is that one can easily miss or misjudge his own purposes, values, and rhetorical strategies.[175] Smith was in fact an ardent defender of commercial society in comparison to actually existing systems exactly because of his commitment to social justice and liberty. He did indeed use a four stages of history scheme to support his claims about the relative benefits of commercial society to other existing socio-economic systems, particularly to argue that commercial society better promoted the material improvement of the poor (and the wealth of society more generally) and the liberty of individuals (since individuals enmeshed in extended networks of interdependency had much more freedom than in the closed dependency relationships associated for example with feudal systems).

But Smith was no simple minded apologist for commerce in general, let alone the mercantilist commercial society of his time. One has to distinguish his descriptive project analysing the relative benefits and costs of existing commercial society from his critical project directed to the improvement of commercial society in the direction of a 'natural system of liberty', focussing on the efforts in ethical and institutional development required to make commercial society a good society (Rasmussen 2008; Hanley 2009; Hanley 2008). When one takes his critical project seriously, it is clear that Smith does not support commercial society for its own sake, but only to the extent that it supports liberty, justice, ethical development, and the poor.

Smith is no hypocrite about these commitments. He argues in favour of general principles of justice at the heart of his political-economics, for example his natural law arguments for individuals' ownership of their own labour as the "foundation of all other property [rights]" (WN I.10.67) clearly imply the injustice of slavery and its inconsistency with property rights in general, as well as supporting the employment freedom of the poor. He also provides an extensive political analysis of the institution of slavery itself in which he highlights

book about political economy in general rather than only about slavery (and anyway in this particular matter the state of British law on the matter was still in flux (LJ 456, editors' fn 71).

[175] For example, Brown criticises the false intimacy implied by Smith's butcher, baker, brewer story in contrast to how commerce 'really worked'. In doing so however his view of Smith seems twice distorted, firstly by considering this story in isolation from and as summing up Smith's whole enterprise (as popularised by some neoclassical 'historians'), and secondly through the lens of what the story says about slavery.

the relationship between the rule of law associated with commercial society and the worsening situation of slaves whose status as property is thus confirmed and locked in. Only where government is not so constrained by the rule of law, and thus able to act "arbitrarily", is it able to intervene with regard to the status and treatment of slaves (WN IV.7.76-77). Here Smith also highlights the role of republican or democratic freedom – if laws are made by the free who are also slave owners, slaves have even less chance of ever securing freedom or better conditions (LJ 181); and the role of wealth which creates a social distance between master and slave that encourages the greatest ill-treatment of slaves (LJ 185).

Far from ignoring contradictions between the benefits of commercial success and the evils of slavery, Smith was fully aware of it and even suggested that such a commercial society was not worth having.

> Opulence and freedom, the two greatest blessings men can possess, tend greatly to the misery of this body of men [slaves], which in most countries where slavery is allowed makes up by far the greatest part. A humane man would wish therefore if slavery has to be generally established that these greatest blessing(s), being incompatible with the happiness of the greatest part of mankind, were never to take place. (LJ 185)

So far from supporting a "dissociative" property rights economics at all costs, Smith clearly, repeatedly, and fervently argued for a better freer and more just society for which property rights were most important as a means and not an end. Far from possessing a "blind optimism" about the benefits of commercial society, Smith took seriously and carefully responded to Rousseau's foundational critiques of commercial society (See for example Rasmussen 2008). Smith was thus a true 'friend of commerce', supporting the project because of its potential virtues, but constructively critical about the shortcomings of the mercantilist society he lived in and commerce in general, providing detailed programmes for institutional correctives (such as universal public education) and ethical development (Hanley 2009; Hanley 2008).

Concluding remarks

Some of the claims that Brown makes against Smithian economics, such as the lack of context and ethical dimension in economic analysis, would seem to have much more bite against neoclassical economic thinking. It is indeed unfortunate that so many critics of contemporary economics, including apparently Brown himself, appear to conflate the classical economics of Smith with the neoclassical economics that has been in the ascendant in the last few decades. While neoclassical economists often lay claim to Smith's authority and legacy,[176] by for example talking up the invisible hand, there are enormous differences and discontinuities in these really quite separate schools and we should not allow the homunculus produced by neoclassical alchemists to replace the real Smith. Indeed the real Smith would appear to offer positive resources for doing a better kind of economics. (I myself have been particularly impressed by the use Amartya Sen has made of the whole Smith - political economist and moral philosopher (Sen 2009; Sen 2010)). I hope to have demonstrated with this rather exegetical analysis that Adam Smith is not the bogeyman he has been presented as. Not only with regard to this single issue of slavery, but in his wider work, Smith was exactly the kind of ethically engaged economist who should be rescued from the straitjacket that partial interpretations have tied him in. He has much more to offer.

[176] According to Ronald Meek, George Stigler started his banquet speech at the bicentennial of the original publication of the Wealth of Nations, "I bring you greetings from Adam Smith, who is alive and well and living in Chicago" (Meek 1977, 3). One need not take such claims at their face value.

Bibliography

Brown, Marvin. 2010. Free enterprise and the economics of slavery. *Real-World Economics Review*, no. 52: 28-39.

Griswold, Charles L. 1999. *Adam Smith and the virtues of enlightenment*. Cambridge University Press.

Hanley, Ryan. 2009. *Adam Smith and the character of virtue*. New York: Cambridge University Press.

Hanley, Ryan Patrick. 2008. Commerce and Corruption: Rousseau's Diagnosis and Adam Smith's Cure. *European Journal of Political Theory* 7, no. 2 (April 1): 137-158. doi: 10.1177/1474885107086445.

Meek, Ronald L. 1977. *Smith, Marx, & after*. Chapman & Hall.

Rasmussen, Dennis Carl. 2008. *The problems and promise of commercial society*. Penn State Press.

Sen, Amartya. 2009. *The Idea of Justice*. Harvard University Press, November 15.

Sen, Amartya. 2010. Adam Smith and the contemporary world. *Erasmus Journal for Philosophy and Economics* 3, no. 1.

Smith, Adam. 1975 [1759, 1790]. *The theory of moral sentiments*, eds. D. D. Raphael, and A. L. Macfie. Oxford: Clarendon Press.

Smith, Adam. 1976 [1776, 1789]. *An inquiry into the nature and causes of the wealth of nations*, eds. R. H. Campbell, A. S. Skinner and W. B. Todd. Oxford: Clarendon Press.

Smith, Adam. 1982 [1762-3, 1766]. *Lectures on jurisprudence*, eds. R. L. Meek, D. D. Raphael, and P. G. Stein. Indianapolis: Liberty Press.

SUGGESTED CITATION:
Thomas Wells, "Adam Smith's real views on slavery: a reply to Marvin Brown", *real-world economics review*, issue no. 53, 26 June 2010, pp. 156-160, http://www.paecon.net/PAEReview/issue53/Wells53.pdf

Comments and reply on Peter Radford's
"Whither economics? What do we tell the students?"

Steve Marglin [Harvard University, USA]

Dear Mr Radford,

I agree with most of what you say about economics, except that it is not as easy as you suggest to separate the study of economics from the study of economies. Keynes said it very well in the preface to the *General Theory*: the hardest part of coming to his new ideas was getting rid of the old ones. The problem is that one needs some kind of framework for studying economies and is thus plunged willy nilly into the study of economics.

Peter Radford

Steve,

Thanks for the comments.

I don't mean to overstate the ease of separation, so I thank you for pointing that out. Clearly any practical work in economics is going to rely heavily on a theoretical framework. My point is that the subject has progressively divorced itself from reality in order to deal with theoretical problems and seems, in my opinion, to have lost its connection with practicality. Subsequent applications of theory are then suspect - the restrictions necessary in theory render its constructs of dubious value in the 'real world'.

I thank you for the reference to Keynes. He could not be more correct. That so much of the internal discourse in economics is focused on the efficacy of his theory goes to show just how tough it is to get rid of those old ideas!

Edward Fullbrook

Steve Marglin, it seems to me, has raised a matter of great importance and one traditionally overlooked by economists at odds with the modern neoclassical mainstream. Dissent generally aims at specific points in the traditional framework and in total may include most or all of its structural timbers. But a new framework, or what Radford calls "coherence", for studying economics never emerges as an alternative. Here work cries out to be done. Without it the revolution will never take place, because either, as Marglin notes, one flounders "wily nilly" or, as Daniel Kahneman has explained, reforms are "noncumulative".

Drawing on his experience with behavioural economics, Kahneman [2003] elucidates the process by which the neoclassical framework – or "cornerstones" as he metaphorizes it – will, if not superseded, silently nullify all reforms even after they have been accepted as valid.

> My first exposure to the psychological assumptions of economics was in a report that Bruno Frey wrote on that subject in the early 1970's. Its first or second sentence stated that the agent of economic theory is rational and selfish, and that his tastes do not change. I found this list quite startling, because I had been professionally trained as a psychologist not to believe a word of it. The gap between the assumptions of our disciplines appeared very large indeed.
>
> Has the gap been narrowed in the intervening 30 years? A search through some introductory textbooks in economics indicates that if there has been any change, it has not yet filtered down to that level: the same assumptions are still in place as the cornerstones of economic analysis. [p. 162]

Kahneman then goes on to explain the mechanism by which the "church" or "theoretical framework" can forever and ever successfully defend itself against any and all significant reform.

> The church of economics has admitted and even rewarded some scholars who would have been considered heretics in earlier periods . . . However, the analytical

> methodology of economics is stable. . . . the standard assumptions about the economic agent are in economic theory for a reason: they allow for tractable analysis. The constraint of tractability can be satisfied with somewhat more complex models, but the number of parameters that can be added is small. One consequence is that the models of behavioral economics cannot stray too far from the original set of assumptions. Another consequence is that theoretical innovations in behavioral economics may be destined to be noncumulative: when a new model is developed to account for an anomaly of the basic theory, the parameters that were modified in earlier models will often be restored to their original settings. [pp. 165-6]

It is worth recalling how the last revolution in economics came about. It took place not through theory but through pedagogy. If Samuelson had any claim to genius it was that he understood better than anyone else that nothing in economics is nearly as important as Economics 101. Marshall, Samuelson's target, understood it also.

Prior to Samuelson's book, Alfred Marshall's *Principles of Economics* (1st edition, 1890) had, beginning with its first edition in 1890, been the leading prototype for introductory economics textbooks. Its opening sentence offered a definition of the subject which underpinned the basic narrative that developed through the long book.

> 'Political Economy or Economics is the study of mankind in the ordinary business of life; it examines that part of individual and social action which is most closely connected with the attainment and with the use of the material requisites of wellbeing.' (8th edition, 1938)

In 1932 Lionel Robbins in his *Essay on the Nature and Significance of Economic Science* in effect redefined economics via a set of axioms declared self-evident and *thereby beyond the empirical realm*, as the "science" of individual choice, the individual being conceived atomisticly, that is, as a determinant self-contained entity. Robbins' redefinition would have counted for little if Samuelson had not then adopted it along with its implicit scientism in his textbook, using it as the reference point from which he constructed a new narrative for academic economics, and with the result that it became and remains the standard narrative approach or "framework" or "church" hymnbook in economics. Because it is inculcated into the student's mind from the first week and every week thereafter, many economists, including heterodox ones, seem unaware of its hold on and significance for their thought. It is this integrated set of empirically false ideas – that economics is the "science" of the choices of isolated individuals with fixed and quantified preferences on a planet which is a subset of the economy – that pervades the economist's mindset, enables economics to proceed in a formalistic manner and justifies to itself ignoring economic phenomena that do not fit its methodology and narrow agenda.

But it does not have to be like this. It is possible to invent for our time an approach to the study of economics that rescues us from willy nilly, that does not disdain empiricism and that restores the dismal science to a cumulative pursuit.

Reference

Kahneman, Daniel, 2003, "A Psychological Perspective on Economics", *AEA Papers and Proceedings*, May 2003, pp. 162-168.

See also

Peter Radford at http://rwer.wordpress.com/2010/06/21/why-%e2%80%9cheterodox-economics%e2%80%9d/#comment-1628 and at http://rwer.wordpress.com/2010/01/25/soul-searching-in-economics/#comment-488

Edward Fullbrook, "The Glass Wall "at http://rwer.wordpress.com/2010/06/25/the-glass-wall/

SUGGESTED CITATION:
Steve Marglin, Peter Radford and Edward Fullbrook, "Comments on Radford", *real-world economics review*, issue no. 53, 26 June 2010, pp. 161-162, http://www.paecon.net/PAEReview/issue53/MarglinRadfordFullbrook53.pdf

ISSN 1755-9472

Articles, comments on and proposals for should be sent to the editor at
pae_news@btinternet.com

Subscriptions to this email journal are free.

Back issues of this journal and other material related to the PAE movement are available at www.paecon.net.

To subscribe to this journal go **here** (http://www.feedblitz.com/f/f.fbz?Sub=332386)

To unsubscribe to this journal go to
http://www.feedblitz.com/f/f.fbz?EmailRemove=_NDUzNDkzMHwzMzIzODZ8cGFlX25ld3NAYnRpbnRlcm5ldC5jb218MTA5OTU4_