

The Review of Austrian Economics

Volume 9,
Numbers 1 and 2, 1995

Murray N. Rothbard

THE | **REVIEW OF AUSTRIAN**
ECONOMICS



Volume 9, Number 1

The Review of Austrian Economics

Founding Editor: Murray N. Rothbard (1926–1995)

Editors

Walter Block

College of the Holy Cross

Hans-Hermann Hoppe

University of Nevada, Las Vegas

Joseph T. Salerno

Pace University

Book Review Editor

David Gordon

Ludwig von Mises Institute

Associate Editors

Steve H. Hanke

The Johns Hopkins University

Jeffrey M. Herbener

Washington and Jefferson College

Managing Editor

Judith F. Thommesen

Editorial Assistant

John S. Thompson

Editorial Board

Martin Anderson

Hoover Institution

Stanford University

Dominick T. Armentano

University of Hartford

Roger Arnold

California State

University,

San Marcos

Charles W. Baird

California State

University, Hayward

Don Bellante

University of South

Florida

James T. Bennett

George Mason

University

Olafur Björnsson

University of Iceland

Samuel Bostaph

University of Dallas

J. B. Bracewell-Milnes

Surrey, England

Yale Brozen

University of Chicago

Alfred G. Cuzan

University of West

Florida

Thomas J. DiLorenzo

Loyola College

Richard M. Ebeling

Hillsdale College

Robert B. Ekelund, Jr.

Auburn University

Joseph Fuhrig

Golden Gate

University

Lowell Galloway

Ohio University

Fred R. Glahe

University of Colorado

David Gordon

Ludwig von Mises

Institute

Robert F. Hébert

Auburn University

Robert Higgs

Seattle University

Ole-Jacob Hoff

Tjøme, Norway

Randall G. Holcombe

Florida State

University

Hans-Hermann Hoppe

University of Nevada

Las Vegas

Edward Kaplan

Western Washington

University

Lord Harris of High Cross

Institute of Economic

Affairs, London

Kurt Leube

Hoover Institution

Stanford University

Stephen O. Littlechild

University of Birmingham

England

Naomi Moldofsky

University of Melbourne

Australia

Thomas Gale Moore

Hoover Institution

Stanford University

John C. Moorhouse

Wake Forest University

David O'Mahoney

University College

Cork, Ireland

E. C. Pasour, Jr.

North Carolina State

University

William H. Peterson

Washington, D.C.

W. Duncan Reekie

University of the

Witwatersrand

South Africa

Morgan O. Reynolds

Joint Economic Committee

Charles K. Rowley

George Mason University

Joseph T. Salerno

Pace University

Arthur Seldon

Institute of Economic

Affairs, London

Sudha Shenoy

University of Newcastle

Australia

Mark Skousen

Rollins College

Gene Smiley

Marquette University

John W. Sommer

Urban Institute

University of North

Carolina, Charlotte

T. Alexander Smith

University of Tennessee

Thomas C. Taylor

Wake Forest University

Richard K. Vedder

Ohio University

Leland B. Yeager

Auburn University

Albert H. Zlabinger

Carl Menger Institute

Vienna

THE REVIEW OF AUSTRIAN ECONOMICS

Volume 9, Number 1

Walter Block, Co-Editor

College of the Holy Cross

Hans-Hermann Hoppe, Co-Editor

University of Nevada, Las Vegas

Joseph T. Salerno, Co-Editor

Pace University

David Gordon, Book Review Editor

Ludwig von Mises Institute

LUDWIG
VON
MISES
INSTITUTE

Auburn University



Kluwer Academic Publishers
Boston/Dordrecht/London

The Review of Austrian Economics

SUBSCRIPTION RATES

The subscription price for Volume 9 (2 issues) is:

For institutions Dfl. 289.00/US \$165.00 including postage and handling.

ORDERING INFORMATION / SAMPLE COPIES

Subscription orders and requests for sample copies should be sent to:

Kluwer Academic Publishers Group
P.O. Box 322
3300 AH Dordrecht
The Netherlands

or

Kluwer Academic Publishers
101 Philip Drive
Assinippi Park
Norwell, MA 02061, USA

or to any subscription agent

All requests for further information concerning *The Review of Austrian Economics* should be sent to:

Kluwer Academic Publishers
101 Philip Drive
Assinippi Park
Norwell, MA 02061, USA

or

Kluwer Academic Publishers
P.O. Box 163
3300 AD Dordrecht
The Netherlands

Photocopying. *In the U.S.A.:* This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Kluwer Academic Publishers for users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the base fee of \$8.00 per copy per article is paid directly to CCC. For those organizations that have been granted a photocopying license by CCC, a separate system of payment has been arranged. The fee code for users of the Transactional Reporting Service is 0889-3047/95/\$8.00.

Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale.

In the rest of the world: Permission to photocopy must be obtained from the copyright owner.

Copyright © 1995 by The Ludwig von Mises Institute.

ISSN 0889-3047

Printed on acid-free paper.

The Review of Austrian Economics has two broadly conceived objectives: (1) to promote the development and extension of Austrian economics and (2) to promote the analysis of contemporary issues in the mainstream of economics from an Austrian perspective.

Correspondence relating to submissions, articles, book reviews, permission to quote, and back issues through volume 3 should be addressed to: *Review of Austrian Economics*, Ludwig von Mises Institute, Auburn University, Auburn, Alabama 36849.

Papers submitted for consideration should be double-spaced, include three copies of the manuscript, and an abstract of not more than 250 words to assist the referees. There is no submission fee. Articles that are accepted will be requested on an IBM format diskette, when possible.

The opinions expressed in this volume by the authors in their respective articles are their own and *The Review of Austrian Economics* assumes no responsibility for these views.

This journal is abstracted and/or indexed in *Sociological Abstracts and Social Planning/Policy & Development Abstracts* and IBZ.

The Ludwig von Mises Institute gratefully acknowledges the generosity of its Members, who made the publication of this journal possible. In particular, it wishes to thank the following Patrons:

Mark M. Adamo	Robert H. Kriebel
Morgan Adams, Jr.	Hugh E. Ledbetter
The Adams Fund	Winston Lehman
O. P. Alford, III	Arthur L. Loeb
Anonymous (3)	Jack E. Magoulakis
George E. Atkinson, Jr.	Dr. Walter Marcyan
Donald Bently	Arthur Marron
John Hamilton Bolstad	Forrest E. Mars, Sr.
Philip G. Brumder	William W. Massey, Jr.
Robert H. Clarke	Roy G. Michell, Jr.
Douglas J. Cook	Roy G. Michell Foundation
Cook Brothers Manufacturing	Frank and Floralie Millsaps
Mr. and Mrs. William C. Daywitt	M. K. Morse
Maino and Dorothy des Granges	The Nehemiah Foundation
Richard A. Elwyn	Mr. and Mrs. Brantley I. Newsom
Dr. Larry J. Eshelman	Ellis L. Phillips, Jr.
James and Vickie Feutz	Clifton B. Rodes
Willard Fischer	James M. Rodney
James W. Frevert	Catherine Dixon Roland
John F. Gemmill	Sheldon Rose
Henry Getz	E. D. Shaw, Jr.
Hager Lumber Company	Quinten E. Ward
Frank W. Heemstra	David and Barbara Westrate
In memory of	Jerald N. Wigutow
John A. and Erlene Hendrickson	Thomas G. Winter
Albert L. Hillman, Jr.	John R. Wright
John Hotonian	Dr. Carl E. Youngstrom
Hugh P. King	

THE | REVIEW OF AUSTRIAN ECONOMICS

Articles

Jörg Guido Hülsmann Free Banking and the Free Bankers	3
Salim Rashid and Abdus Samad Portfolio Management of the Free Banks of Illinois: An Examination of Historical Allegations	55
Walter Block and Kenneth M. Garschina Hayek, Business Cycles and Fractional Reserve Banking: Continuing the De-Homogenization Process	77
Pascal Salin The Myth of the Income Effect	95
Anthony de Jasay Hayek: Some Missing Pieces	107
David W. Boyd Vertical Restraints and the Retail Free Riding Problem: An Austrian Perspective	119

Notes and Replies

Leland B. Yeager Rejoinder: Salerno on Calculation, Knowledge, and Appraisalment	137
Joseph T. Salerno A Final Word: Calculation, Knowledge, and Appraisalment	141
Hans-Hermann Hoppe Socialism: A Property or Knowledge Problem?	143
Jeffrey M. Herbener Calculation and the Question of Arithmetic	151

Review Essays

Roger W. Garrison Keynes Was a Keynesian	165
Murray N. Rothbard Intimidation by Rhetoric	173

Book Review

Murray N. Rothbard, <i>Economic Thought Before Adam Smith</i> (vol. I) and <i>Classical Economics</i> (vol. II) Reviewed by Leland B. Yeager	181
--	-----

Articles

Free Banking and the Free Bankers

Jörg Guido Hülsmann

The literature on free banking has expanded dramatically in the last two decades. A young generation of economists has regained interest in questions of money, banking, and currency that, for a very long time, had disappeared from broad discussion. This renewed interest was partly sparked by poor results from government regulation of the money supply by central banks, as well as other legal devices and restrictions. Such failures have undermined the once-common belief that blessings can flow from government monetary meddling. Because free banking was the historical predecessor of and natural alternative to monetary interventions, the theory and practice of free banking has attracted a great deal of interest.

It is common for people eager to fight for a specific cause to employ intellectual means unfit to serve their ends. As a result, they may achieve the opposite of their intentions, undermining the ideals and ideas they are seeking to promote.

Such is the case with free banking. The case for authentic free banking has been obscured by the strongest defenders of free banking.¹ In defending views that are not only unrelated to free banking but even fallacious, the free bankers do much harm to their case, inadvertently adding weight to the critique of free banking offered by advocates of central banking and government money.

Jörg Guido Hülsmann is instructor of economics at the Technische Akademie Wuppertal. I wish to thank the referees for extensive comments on this paper.

¹Kevin Dowd, David Glasner, Steven Horwitz, A. J. Rolnick, Larry Sechrest, George Selgin, Lawrence White, and Richard Timberlake. I shall concentrate my discussion on the works of Dowd, Selgin, and White. It is here that the doctrine is elaborated. By contrast, the contribution of Sechrest consists of a formal, i.e., mathematical, expression of their tenets; Glasner and Horwitz base their works heavily on Selgin's *Theory of Free Banking*; and Rolnick and Timberlake have contributed applications of free banking theory to historical episodes.

The Review of Austrian Economics Vol. 9, No. 1 (1996): 3–53
ISSN 0889-3047

We can divide the advocates of free banking into two groups. The first group proceeds from the assumption that the money and banking sector can operate with virtually no money at all. Within this group, there are additional disagreements. They concern the questions of whether *laissez-faire* would only be efficient in a situation in which no money is used² or whether free banking would even actively bring such a situation about.³

The internal dispute within this first group is not essential to understanding a more fundamental fallacy of its theory. The hope of a high degree of division of labor without the use of money is futile; there can be no "unit of account" without indirect exchange. Economic calculation presupposes the use of a general medium of exchange. Everyone is, indeed, free to *translate* a money calculation into whatever unit he likes.

Using, for example, coat hangers as the "unit of account," one could calculate a profit of 1000 coat hangers from an investment. Yet this calculation is nothing but an algebraic expression of: "*For the money* which was spent in the investment I could have bought 500 coat hangers, and *for the money* I received in exchange of the product of the investment I could buy 1500 coat hangers. If after my investment, my money can buy more coat hangers, I am richer than before. In the same sense my investment can be regarded as profitable." No *numéraire* or "commodity bundle" or anything else could serve as a calculation unit if there were no money in use. No indirect exchange can be settled without the use of money.

The focus of our thesis, therefore, lies on the discussion of the second group, comprising the more common free bankers. This group shares the view that no modern society is possible without the use of money.⁴ They disagree, however, over the social and

²See Fischer Black, "Banking and Interest Rates in a World Without Money," *Journal of Banking Research* 1 (1970); Eugene F. Fama, "Banking in a Theory of Finance," *Journal of Monetary Economics* 15 (1980); Robert E. Hall, *Inflation, Causes and Effects* (Chicago: University of Chicago Press, 1982); Robert L. Greenfield and Leland B. Yeager, "Laissez-faire Approach to Monetary Stability," *Journal of Money, Credit, and Banking* 15 (1983).

³See Neil Wallace, "A Legal Restrictions Theory of the Demand for 'Money' and the Role of Monetary Policy," *Federal Reserve Bank of Minneapolis Quarterly Review* (1983); Thomas Sargent and Neil Wallace, "The Real-Bills Doctrine versus the Quantity Theory: A Reconsideration," *Journal of Political Economy* 90 (1982).

⁴With the exception of Kevin Dowd, see, *The State and the Monetary System* (London: Phillip Allan, 1989), p. 188; idem., *Laissez-faire Banking* (New York: Routledge, 1993), pp. 66f; David Glasner, *Free Banking and Monetary Reform* (New York: Cambridge University Press, 1989), p. 240f; and Richard Timberlake, *Gold, Greenbacks, and the Constitution* (Berryville, Va.: George Edward Durrell

economic effects caused by money substitutes. Some of them claim that the practicability of free banking requires full coverage of money substitutes. If the issuer of each ticket grants the right to redeem, at par and at the arbitrary request of the holder, a certain quantity of money has to be held as a 100 percent reserve. Where this is the case, the money substitutes have the character of certificates. By contrast, tickets issued on a less than 100 percent reserve are called fiduciary money substitutes. They are no longer certificates because they are only fractionally covered by the quantity of money to which they represent a claim.

In addition, the “free bankers” of this second group claim that fractional reserve banking would not only be practicable but also beneficial. Predictably, then, they also argue that 100 percent reserve banking has considerable disadvantages. A critique of their tenets, therefore, has to embrace both their arguments in favor of fractional and against 100 percent reserves. First I will discuss free banking on a 100 percent gold standard and the principal arguments that have been pronounced against it. Then I turn to the alleged benefits of free banking on a fractional reserve basis. Finally, I try to explain why neither fractional banking nor banking on a fiat money base can be practicable.

Free Banking on a 100 Percent Gold Standard

Money and Substitutes for Money

In monetary theory, there is hardly a word (apart from inflation) that causes as much confusion as the word money itself. It is vital to distinguish money from money substitutes. Yet this distinction is obfuscated by calling the latter “inside” money and the former “outside” money. The same confusion results from spurious talk of “base money,” “basic money,” or “high-powered money.” These terms suggest that there is no practical difference between them; all the instruments in question are somehow “money.” The climax of all this rhetorical excess is undoubtedly attained when fractional reserve advocates George Selgin and Lawrence White speak of gold or the gold dollar “as a substitute for bank deposits.”⁵

Foundation, 1991), pp. 60ff. For a critique of the latter see Rothbard, “Aurophobia: or, Free Banking on What Standard?,” *Review of Austrian Economics* 6, no. 1 (1992): 97–108.

⁵George Selgin and Lawrence White, “How Would the Invisible Hand Handle Money?,” *Journal of Economic Literature* 32 (1994): 1737. See also Lawrence White, “Identifying Money,” in his *Competition and Currency* (New York: New York University Press, 1989), pp. 206ff.

Does there exist something like a second kind of money? Imagine two scenarios faced by moviegoers. In the first, an individual purchases a ticket, but before entering the theater and taking his seat he decides not to see the movie because there are more urgent things to do. He therefore sells it to somebody who does not yet have one. In the second scenario, the same individual enters the cinema, redeems his ticket with the usher, and takes his seat, but then decides not to stay. His neighbor has found a friend who wants to sit in his place, and he sells his seat to him.

Clearly, in this second scenario his neighbor has not purchased a *substitute for a movie ticket*. He has purchased the seat beside him for the time that the movie is shown. The same holds true in the first case. The first moviegoer did not sell a piece of paper; but sold a seat to a certain showing of the movie. Otherwise he would not have been able to get something in exchange for the ticket. Nobody interested in seeing a film would buy sheets of paper called "tickets" if they were not a means for seeing the film. Neither is anybody eager to buy sheets of paper called banknotes were it not for the convenient disposition of money.

It is also problematic to describe the relationship between money and money substitutes as one of fixed parity or convertibility. In a larger sense all goods exchanged against one another have a parity, that is, the exchange rate. In the same sense, all goods exchanged on the market have proven to be convertible into one another. However, this does not mean that the parity is already implied in the existence of the exchanged goods.

Assume that Paul exchanges eight hours of his work against one ounce of gold. *After* the exchange has taken place, one can say that Paul's work has been converted into gold or that gold has been converted into Paul's work. Yet the existence of one ounce of gold does not imply that one will receive eight hours of Paul's work for it. Nor does Paul's capacity to work stem from the fact that it can possibly be exchanged against one ounce of gold. The existence of the gold and the existence of Paul's capacity to work are independent from each other. Their exchange rate is not implied in their mere existence.

It is different with money substitutes. They can only come into being as a claim, a part of a contract, that fixes their exchange rate to money. They are signs, expressions for the disposition of a certain quantity of money. When they are exchanged against money they are redeemed. Redeemability is the original meaning of the term convertibility. A document that is convertible in this sense can never have a value different from the object that it gives

a claim to. A convertible currency—money substitutes in the form of bank notes—can neither be a money nor a standard.⁶ Only irredeemable notes are money—that is, fiat money. They are valued separately because they can be used independently from other goods.

Banknotes and demand deposits are money in only one case: if they do not represent claims. Obviously such a situation cannot come about unless the redemption promise is broken. Breaking a contract amounts to an expropriation of the partner in exchange. That our present money consists of irredeemable banknotes and demand deposits—of central banks—is the result of government-initiated expropriations of money that characterize modern history.⁷ Banknotes can only be government (fiat) money because no other agent in a modern state can break contracts on such a wide scale without fear of punishment.

In a system of free banking—whether on a fractional or 100-percent-reserve basis—the demand deposits and banknotes of the competing banks are *substitutes*. They represent a convenient means of documenting claims on money. In exchanging these tickets, one exchanges ultimately (presently existing) money of which they are considered to be representative. Under a gold standard, the exchange of banknotes signifies the exchange of weights of gold. Tickets and other signs are useful because they are not as heavy or voluminous as the goods that are the real objects of the exchange.

*Would there be Money Substitutes
on a 100 Percent Reserve Basis?*

Under a 100 percent gold standard all money substitutes are entirely covered by gold. For each checking account and for each banknote held by the public, the designated amount of gold lies in the vaults of some bank.⁸ The banks do not lend this gold to other market participants. They hold it and permit the owner to use some substitutes for his gold that facilitate his market exchanges.

⁶See the contrary opinion of White, *ibid.*, pp. 134f.

⁷See, for example, V. C. Smith, *The Rationale of Central Banking* (1936; Indianapolis, Ind.: Liberty Classics, 1990); Kevin Dowd, *Laissez-faire Banking*, esp. chap. 10.

⁸It is noteworthy that what is said about banks applies to virtually all financial intermediaries dealing with money substitutes. Money substitutes are not only banknotes and demand deposits but principally *all* claims that have to be redeemed at par into money whenever the holder of the claim likes to have money substitutes. See in particular Murray N. Rothbard's excellent analysis of money substitutes in the 1920s in *America's Great Depression*, 4th ed. (New York: Richardson and Snyder, 1983), p. 83.

In dealing with demand deposits and notes, banks do not act as financial intermediaries but as warehouses. Financial intermediation, then, can only be provided if and insofar as market participants temporarily *renounce* a claim to the disposition of their money and give it into the disposition of their banks. This is the meaning of the term credit. Under 100 percent reserve banking, credit given by money owners is the necessary condition of financial intermediation. Only if a gold owner has lent his gold to his bank can the bank, in turn, lend this gold to other market participants. Banks are thus engaged in two completely distinct businesses. On the one hand, there is the warehouse business with money substitutes; on the other hand, there is the credit business with money that has been given for their exclusive disposition. There is no reason to assume that these two businesses must always be performed by the same company. Specialization can lead to exclusive gold warehouses and exclusive financial intermediaries.⁹

The money owners profit from the use of banknotes and checking accounts. They do not have to charge themselves with the inconveniences that go hand-in-hand with the use of relatively voluminous and heavy metallic money. In the case of checking accounts they can also avoid the risks of keeping their money at their homes, for no check is valid without their signature.

The holders of demand deposits, in one way or another, have to pay for these services. Their bank will have to charge them with the full costs of security provisions, and of the factors of production the bank has to buy in order to deal with depositors. Otherwise, either the bank's profits would be reduced, or it would have to charge the costs to its financial intermediation business. In the latter case, the bank would become less attractive in comparison to its competitors. It would either have to charge higher interest rates for the money it lends or pay lower interest rates on the money it borrows.

One hundred percent reserve banking differs from banking as we know it from our daily transactions because interest could no longer be paid *on* demand deposits, but a fee would have to be paid *for* them. It is therefore very probable that, should such a system be introduced, fewer people than today would like to hold their money with the banks and use money substitutes instead.

⁹See Condé Raguet's discussion "Of Banks of Deposits, Banks of Discount, and Banks of Circulation" in his *Treatise on Currency and Banking* (New York, 1840), pp. 67ff.

One cannot say that no money owner would accept such a deal. Questions of this kind can only be answered empirically, that is, not before banks and their customers actually deal with such a situation. If there is at least one customer to whom using money substitutes is more important than the fee due, then there will be money substitutes on a 100 percent reserve basis.

Warehouses for money would not be more unusual than warehouses for other commodities. Considering the conveniences linked to the use of money substitutes, there are good reasons to believe that the latter will find employment especially in the performance of big payments. Yet all other transactions will largely be dominated by specie. Thus, under 100 percent reserve banking, gold would certainly not be outcompeted by its substitutes. It would always stay in circulation.¹⁰ However, in a big and growing market, the inconveniences linked to the use of relatively heavy and voluminous gold (and especially silver) money would be progressively reduced. The more transactions are effected on the market, the more purchasing power would accrue to a given quantity of gold.¹¹

The Consequences of Individual Failure Under 100 Percent Reserve Banking

Under 100 percent reserve banking all banks can operate independently of one another. The illiquidity of one bank never implies the illiquidity of the others. If one bank is becoming illiquid, it is forced to retire immediately all the money it has lent to other market participants (and, hence, to other banks). However, this will never lead to the illiquidity of those borrowers who have not engaged themselves in the transformation of maturity. Illiquidity will be limited to those borrowers who had put the borrowed money into employments that are more lengthy than the credit term and who now are unable to meet their obligations.

At all times and in all places there will be market participants whose speculations prove to be erroneous and who fail to fulfill their contracts. Such failure always has negative repercussions

¹⁰See, Hans-Hermann Hoppe, "How is Fiat Money Possible," *Review of Austrian Economics* 7, no. 2 (1995): 57.

¹¹This argument was used by Condillac in order to claim that not only the quantity of money is rather irrelevant but that, on the contrary, it would be advantageous if it were smaller ("On voit donc qu'il est assez indifférent qu'il y ait beaucoup d'argent, et qu'il serait même avantageux qu'il y en eût moins. En effet, le commerce se ferait plus commodément. Quel embarras ne serait-ce pas si l'argent était aussi commun que le fer?" *Le Commerce et le gouvernement* (Paris, 1795), p. 87).

on their business partners and regularly leads to the failure of some of them, too. But so long as error is limited to only a few market participants, it cannot have, under 100 percent reserve banking, repercussions on the whole economy. Ruins will then always impede only a very limited group of enterprises. There will always be a problem concerning the immediate business environment of the errant. No central bank is needed to limit their repercussions further.

Nobody has ever raised the objection that a 100 percent reserve system would lead to wide-spread business failure. Nobody has ever been able to prove that this system cannot endure, that it must inevitably lead to its own destruction. All of its critics have pointed to some alleged shortcomings of 100 percent reserve banking for which they propose fractional reserves as an antidote. In the next section the question of such "shortcomings" will be examined.

Critics of 100 Percent Reserve Banking

The Alleged Costs of 100 Percent Reserve Banking

The most common objection against a full coverage of money substitutes is that the system would be too costly. The money in the vaults of the banks is lying idle. It could be better used for other purposes, for example lending it to someone in need of a credit. This idea is entirely wrong. The confusion that constantly arises about this issue is related to the concept of cost itself. Costs are always the costs of an action that an individual confronts. They consist of all the desirable effects that, in the eyes of the actor, cannot be brought about because he has preferred to aim at some other ends. Costs are the expected forgone utility.

The concept of cost has no meaning whatever apart from choice. It cannot be understood if only one action is considered apart from two alternative actions. Every actor is always confronted with *some* costs. The use of money is no exception. From the point of view of a money user, it is obvious that holding money, whether in cash or in form of a bank account, is costly. Indeed, he could employ it in buying some useful commodity or service. Accordingly, it is also costly for the bank to keep large stocks of money. There are always some people ready to pay at least *some* interest rate on additional funds.

However, does the mere fact that an action is costly represent a shortcoming of this action? Does the mere existence of costs represent a shortcoming of the use of money? Clearly, the answer

is no. Costs are the forgone utility of an action that is not carried out because another action has been preferred. Hence, so long as a person has to choose from among the specified ends the chosen action must be costly. And so must be the use of all other means that could also be used in another way. We can employ no commodity without having it at our discretion, namely, without holding it. Therefore holding it must be costly in one respect or another. There are always costs with the *holding* of money because its *use* implies holding it.

It seems as if at least some of the free bankers agree with this argument. They concede the fact “that the use of money carries with it certain social costs (forgone benefits of barter) does not compel one to conclude that its costs outweigh its benefits.”¹² Indeed, the very use of money *implies* that for its user the benefits outweigh the costs. Money is always used in spite of its costs. However, the free bankers fail to see what this implies about 100 percent reserve. They continue to adhere to the spurious distinction between money hoards and money in circulation. In their eyes, there are people who do not want to hold money but only want to use it in market exchanges. Where no money is held, they suggest, there can be no costs. This reasoning is fallacious. It is impossible “to receive money in exchange for other goods and services” without having the “desire to hold money balances.”¹³ The use of money must always be costly.

True, say some economists, the mere fact that costs are inextricably linked to the employment of all non-specific means cannot reasonably be considered as a disadvantage. But does this compel us to satisfy ourselves with the present level of costs? All great inventions have this in common that they reduce the costs of action. Why, then, should we not seek for such cost reductions in the realm of money?

Look, for example, at cars parked idly in the streets while their users are at work. They just use their cars to drive from their homes to work in the morning and in the evening they drive them back home again. Many more services could be rendered by these cars if their owners would allow other people to use them during their worktime. The same thing holds true for money. Instead of lying idly in the vaults of the banks, it could be usefully employed by other people in the meantime.

¹²Lawrence White, *Competition and Currency*, p. 200.

¹³George Selgin, *The Theory of Free Banking* (Washington, D.C.: Cato Institute, 1988), p. 53.

Thus, a certain quantity of gold could serve several bank customers at the same time. This is the nature of fractional reserve banking.

We do not have to discuss the question if whether cars can render *additional* services. For the sake of the argument, we might admit that. Let us suppose that parked cars could render additional services when they are used—with or without the consent of their owners—by other people. What is at stake is the question of whether the same holds true for money. It is this question, however, that we have to answer in the negative. For the services that stem from the use of a certain quantity of money depend on money prices, and money prices depend on the use of the existing quantity of money. It is by the use of idle money from demand deposits that money prices will unavoidably be enhanced. Thus, not only the owners of the demand deposits that were lent out but *all* owners of money, be it in the form of cash or in the form of money substitutes, will find the purchasing power of their money balances reduced. The use of idle money hoards is paid by the owners of these hoards and all other money owners. No other outcome is conceivable because the mere intensification of the use of money does not imply the intensification of the production of goods. The use of money and of its substitutes is always costly. If it is not the holder who is charged with these costs it must be someone else.

It is true that all new technical devices to economize the use of money have resulted in a tendency to higher money prices. The same will inevitably hold true for all future improvements of this kind and thus they have the same effect as a further reduction of the reserve ratio of money substitutes. But does this prove that there is no other difference between them which is crucial? Does it not simply represent another proof of the virtual irrelevance of the money price level?

There is no need to enter into the discussion about the importance of money prices. We rather have to emphasize the difference between two origins of a more intense use of money. One is entrepreneurial innovation and the other is the reduction of the reserve ratio. The great innovations of banking history such as banknotes, checking accounts, clearing houses, and credit cards have brought advantages for all market participants. They economized factors of production that the banks formerly had employed in the service of their customers. Less money had to be spent in the production of these banking services so more could be spent for other market transactions. The same

effect was caused by all innovations of non-bankers permitting them to keep smaller money balances. New techniques for business accountancy, for the planning of market transactions, and for business forecasts fall in this category.

On the other hand, a smaller reserve fraction merely means inflation, viz., an extension of the quantity of money in the larger sense. Yet, as no factor of employment has been reduced, no additional production can result from it. There can be no doubt that the first mentioned innovations are not inflationary, viz., increase the quantity of money in the larger sense. They lead to a more intense use of the existing quantity of money or, in other terms, enhance the velocity of circulation. This is what causes an increase of money prices on the market. The profits derived from productive innovations are a reward for an achievement that is useful for all market participants. By contrast, profits derived from inflationary reductions of the reserve fraction simply represent fraud. No use of a factor of production has been reduced. The banker gains something which is taken from other people.

Financial Intermediation Under 100 Percent Reserve Banking

Implicit in all arguments against banking on a 100 percent gold standard is the conviction that this system would gravely impede financial intermediation. In the judgment of moderate free bankers, such as Larry Sechrest,

First, with 100 percent reserves, banks cannot make loans from their deposits. Every dollar deposited must be held, ready to be redeemed, at all times. This severely restricts the available credit in the society. One could make a very plausible argument that much of the real economic growth that has occurred would have been impossible in a world of 100 percent reserve banking. Furthermore, banks resent such an imposition.¹⁴

This expression fatally recalls the inflationist real-bills doctrine. Therefore, some free bankers advance a more radical argument. They say that 100 percent reserve banking makes financial intermediation impossible. This is, however, untrue; even if credit were *restricted* by 100 percent reserve banking (which is not the case) there would still be credit in this system. To be sure

¹⁴Larry Sechrest, *Free Banking: Theory, History, and a Laissez-faire Model* (Westport, Conn.: Quorum Books, 1993), p. 66.

there would be no intermediation of demand deposits because the disposition of them would entirely be reserved to the depositors. Yet even in a system of fractional reserve banking the intermediation of demand deposits represents but a part of the whole intermediation business. By far the biggest part of the money lent by the banks has been temporarily given into their *exclusive* disposition. Therefore, the pretension that under 100 percent reserves "banks would be unable to lend"¹⁵ is untenable. One does not have to quarrel about whether the word credit, or the expression "true financial intermediation,"¹⁶ should be reserved for lending operations on the basis of demand deposits (fiduciary money issues). The only relevant issue is whether there is still financial intermediation under 100 percent reserve banking. This cannot be contested.

*The Alleged Dangers of Money Shortages
and of Changes of the Price Level*

The case for fractional reserve banking is entirely based on the age-old equivalence idea. According to this idea each commodity *corresponds* to some quantity of money. The exchange of a bigger quantity of goods on the market is only possible if the quantity of money increases, too. Devastating results could result from a "fear of currency shortage."¹⁷ The prospect of a rigidly limited quantity of money, say the free bankers, could drive the market participants to enhance their money holdings. This would precipitate a real money shortage even if there had been none in the beginning. It is obvious that this argument not only applies to gold but to all other goods as well. The quantities of shoes, bread, and bottles of milk are no less limited than the quantity of money. Nevertheless there are no general fears of shoe shortages. Neither is it necessary to invent special devices to prevent them.

However, this is not the whole of the picture painted by the free bankers. Fractional reserve banking is needed because metallic money cannot increase in a degree sufficient to permit all market exchanges. It is needed to provide "transfer credit."¹⁸

¹⁵Dowd, *The State and the Monetary System*, p. 25.

¹⁶Steven Horwitz, *Monetary Evolution, Free Banking, and Economic Order* (Boulder, Colo.: Westview Press, 1992), p. 115.

¹⁷Selgin and White, "How Would the Invisible Hand Handle Money?": 1726.

¹⁸See Selgin, *The Theory of Free Banking*, pp. 60ff. Transfer credit is "credit granted by banks in recognition of people's desire to abstain from spending by holding balances of inside money" (p. 60). This of course, is no definition. Every use of money implies the holding of it.

Transfer credit, they say, is necessary to prevent disruptive consequences that otherwise would follow. Principally, they say, each increase in the demand for money would cause unfavorable money shortages for it withdraws money from circulation:

Consider what happens when the supply of money fails to increase in response to an increase in demand for money on the part of wage earners. The wage earners attempt to increase their money balances by reducing their purchases of consumer products, but there is no offsetting increase in demand due to increased, bank-financed expenditures. Therefore, the reduction in demand leads to an accumulation of goods inventories. Businesses' nominal revenues become deficient relative to outlays for factors of production—the difference representing the money that wage earners have withdrawn from circulation. Since each entrepreneur notices a deficiency of his own revenues only, without perceiving it as a mere prelude to a general fall in prices *including factor prices*, he views the falling off of demand for his product as symbolizing (at least in part) a lasting decline in the profitability of his particular line of business. If all entrepreneurs reduce their output, the result is a general downturn, which ends only once a general fall in prices raises the real supply of money to its desired level.

As was said previously, such a crisis can occur only if banks fail to respond adequately to a general increase in the demand for inside money.¹⁹

This reasoning is central for the doctrine of fractional reserve banking. There are several fallacies in it. Even if it *were* correct, there would be no way to explain why prices can ever fall. Yet this is what the free bankers consider as the long-run outcome of a growing economy.

Most importantly, the above statement is but half of the story. The other half is the story of wages.²⁰ If an entrepreneur faces reduced demand for his products, he sooner or later has to pay lower wage rates. Now, if a worker accepts this, the output of this enterprise is not reduced. It remains profitable and can stay in business. If a worker does not accept the lower wage rate, he will

¹⁹Ibid., p. 55.

²⁰See the following, for example, W. H. Hutt, *The Keynesian Episode* (Indianapolis: Liberty Press, 1979), pp. 51ff.

sooner or later have to look for another employment, thus reducing wage rates elsewhere. Other businesses that hitherto were submarginal become profitable. In either case there can be no general reduction of output. Wage earners will have lower nominal incomes.

Yet, all other prices are lower, too. Thus their real incomes have not declined. Even if all wage earners decided suddenly to bury their banknotes in pillows or burn them, there would be no need and no possibility to adequately increase the supply of banknotes. To be sure, there would be some disruptive elements in this scenario. Yet, it is not the falling prices that are disruptive, but the general folly that drives *all* market participants to burn their banknotes. Falling prices are nothing but a symptom of an adjustment taking place. Preventing prices from falling amounts to curing the symptom and leaving the disease untouched. General output or aggregate demand can neither be conserved nor enhanced by increasing the money supply. The free bankers have not yet learned the lesson of Say's Law.

Some of the free bankers have filled volumes with studies on the history of banking and, still, are blind to the most important issues of money and banking. With the opponents of gold, they share the conviction that money is only optimal if it is flexibly supplied according to the changing scope of its employment or of needs. There can be no greater fallacy in monetary theory. No issue is more fundamental. Therefore the insight of classical economics has to be repeated again: *The quantity of money is irrelevant for the benefits derived from its use, in the long run and in the short run.* There is no need and no possibility to adjust it according to its changing employment. There is no need because the adjustment can be achieved by a change of prices and particularly a change in wages. But most importantly, there is *no possibility* of an "adjustment" of the quantity of money. Even if one could succeed in replacing the money exactly there where it is "withheld" (which would be close to a miracle) one would need an angel to inform each market participant about the structure of prices that is now likely to be created.²¹

There is no meaningful way to define a demand for money that could exceed the supply of money (the existing money stock). An

²¹See the analogous remarks of Mises concerning the possibility of stabilizing the value of money in his *Theory of Money and Credit* (Irvington-on-Hudson, N.Y.: Foundation for Economic Education, 1971), pp. 123–31.

ever increasing quantity of commodities and services can be sold on the market with one and the same supply of money.²² The argument can be reduced to the conviction that “if prices go up we need more money to sell all the goods.” However, the mere fact that one price or even all prices did already go up with the use of *unchanged* money stock proves that the latter does not have to be increased.

Every existing good can be exchanged on the market. The crucial question is whether the selling prices render its production profitable or not. Unprofitable investments prevent more urgent productions. This is why they are unprofitable. If transfer credit is given to make them profitable, the satisfaction of more urgent wants is artificially prevented.

On Some Alleged Advantages of Fractional Reserve Banking

Is Fractional Reserve Banking the Necessary Outcome of an Unhampered Market?

Fractional reserve banking has been represented as the necessary outcome of an unhampered market.²³ If this were true it would be a strong support for the claims of the free bankers. For whatever was undertaken by any other agent to establish a different system, there would always prevail a tendency toward fractional reserve banking.

It is most convenient to clarify the nature of this argument because some of its advocates believe it to be “causal-genetic,” an expression which Schumpeter used to distinguish Austrian economics from other approaches. An abstract summary of it could run like this: First one points at a problem of action, for example, the problem that “double coincidence” in a barter economy is very rare so that most people willing to sell the goods which they do not need personally would not be able to exchange on the market. Then one shows that this problem can be solved by a certain behavior that was until now unknown.

In our example this would be the invention of indirect exchange: using a medium of exchange, people are no longer dependent on the

²²For example David Ricardo, *The High Price of Bullion*, Works 3 (London: John Murray, 1811), p. 73.

²³G. Selgin and L. H. White, “The Evolution of a Free Banking System,” in Selgin, *The Theory of Free Banking*, chap. 2, and in White, *Competition and Currency*, chap. 12; also Dowd, *Laissez-faire Banking*, pp. 26–33, 59–68.

improbable case of "double coincidence." With this solution of the old problem, however, new problems are arising by which no one has been previously confronted. One of these new problems is linked to economic calculation. Economic calculation cannot be successfully executed without the use of a medium of exchange. The calculated planning of action reaches as far as the price system that is constituted by the use of the particular medium of exchange.

Comparing the prices expressed in a medium of exchange that will probably be realized on the market permits us to evaluate the probable success of even the most complex projects with a hitherto unachievable precision. On the other hand, one of the problems that is linked to economic calculation is the homogeneity of the medium of exchange. If the different items of the total quantity of a medium of exchange are not of a sufficiently homogeneous quality, no calculation can be successfully put into action. A new solution is required to solve the new problem. As should be clear by now, whatever solution will be applied, it will be at the base of other problems that need other solutions, and so forth.

This essay is not concerned with questions of method. Yet, fractional reserve banking is recommended because it allegedly represents the necessary outcome of the operation of the unhampered market which in turn can allegedly be deduced by the above method. The latter, therefore, needs some consideration.

It is very important to realize that in economics there are two types of arguments of which one could say that they feature evolutions. One argument is purely logical. This is, for example, the case for the necessary evolution that we call the business cycle.²⁴ A business cycle takes place after the injection of additional quantities of money through the credit system. Whatever the market participants will do in such a situation, they cannot prevent the additional quantity of money from exercising an additional effect on the price structure. After the injection of new money, many projects seem to be (are calculated to be) profitable that did not seem so before. Projects are started which would not have been started without the injection of new money. Indeed, saying that additional quantities are *injected* into the market through the credit system means *that* they are borrowed. Then the price (the interest rate) must be lower than it would otherwise have been. As this interest

²⁴See Ludwig von Mises, *Human Action* (Chicago: Henry Regnery, 1949), pp. 571ff; and Murray N. Rothbard, *Man, Economy, and State* (Los Angeles: Nash, 1962), pp. 850ff.

rate *cannot* last but *must* go up it represents an *additional* source of error for market participants.²⁵

The alleged deduction in the theory of fractional reserve banking is not of this kind. Essentially it is a historical account, even if it does not feature *our* history. The necessity of the evolution it describes is only an empirical, that is *ex post*, necessity. Of course, we know that in all types of barter societies, the problem of double coincidence exists. We also know that man has discovered indirect exchange. Yet, this invention, as every other invention as well, was in no way inevitable. In all places and at all times action is confronted with problems. Only *ex post* are we often capable of saying if and in how far a certain behavior represents a solution and to what. This is what can be achieved with causal-genetic approaches to the evolution of monetary institutions. And this is *all* they can achieve. They are a kind of very abstract history of monetary institutions, a history of what would have happened if government had not intervened in a misconceived manner.

Now, let us disregard the question of whether it is appropriate or not in this context to neglect government interventions. The only question we have to face is whether there are any problems of action that, *by their pure existence*, imply that a certain solution—indirect exchange, clearing houses, fractional reserve banking, etc.—be invented. Does an empty refrigerator imply that it will be filled? Did gravitation create the relativity-theory to let man fly to the moon? Did the weakness of our eyes invent X-rays to see through a patient's skin? Did idle gold hoards lead to fractional reserve banking? If this were the case, the causal-genetic process would be a sound line of reasoning. Yet, it is not the case. It is undisputed that all the institutions that are allegedly *deduced* from problems represent, in some manner at least, solutions to existing problems. However, this is no proof that other outcomes would not be possible. Fractional reserve banking

²⁵By the way, it is not true that a reduction of the inflated money stock is the cause of crises. It is already the widespread *injection* of additional money via the credit system which implies that money calculation has to fail on a wide scale. Once the failure becomes obvious in the form of a crisis, a reduction of the money stock has the effect of accelerating recovery. Hence, one cannot claim that "Austrian economists such as Rothbard *add* that it was the Fed's expansionary policies during the 1920s that precipitated the crisis, which was exacerbated by the Fed's later inaction" Steven Horwitz, *Monetary Evolution, Free Banking, and Economic Order*, p. 182 [emph. added]. This claim presupposes that monetary expansion is but a detail in the picture of business cycles and that Rothbard shared the view that it is the drop of the money stock which creates crisis. Neither is true.

could be but a part of all the possible solutions. To state a problem and then present one (now known) solution is no proof that the problem already *meant* this solution. If this implication cannot be demonstrated, the argument can never be general, that is, valid for the markets of all times and places. It then refers only to one particular outcome, not to all the outcomes that the unhampered market must take.

The approach championed by the free bankers contains no argument of the kind required to prove that an unhampered market leads to fractional reserve banking. It cannot be claimed in defense of the case for fractional reserves.

Does Fractional Reserve Banking Lead to Monetary Equilibrium?

Free bankers Kevin Dowd and Lawrence White say they do not defend the real-bills doctrine.²⁶ Even so, they have recognized the proximity between their tenets and this fallacious doctrine. Yet all their efforts to distinguish between the two have proved to be futile. There is no difference between a money substitute issued to give a real-bill credit and a money substitute issued to give transfer credit.²⁷ Both are credits out of thin air, that is, no credits at all. Contrary to their pretensions, the free bankers are nothing but the modern advocates of the real-bills doctrine.

There is but one quite modern feature in their argument. It is the conviction that only fractional reserve banking leads to monetary equilibrium. The latter is supposed to be the state of affairs that prevails when "there is neither an excess demand for money nor an excess supply of it at the existing level of prices."²⁸ According to Selgin, the lending process in a fractional-reserve banking system equilibrates money supply and demand because:

Whenever a bank expands its liabilities in the process of making new loans and investments, it is the holders of the liabilities who are the ultimate lenders of credit, and what

²⁶For a vain attempt to prove the contrary, see, Dowd, *The State and the Monetary System*, p. 60ff.

²⁷In a brilliant analysis Fritz Machlup demonstrated that the time horizon for which a credit is given has nothing to do with the time horizon of its employment. It is thus illusory to believe that fiduciary issues would only finance transfers (*Börsenkredit, Industriekredit und Kapitalbildung* [Vienna, 1931], pp. 139, 179ff).

²⁸Selgin, *The Theory of Free Banking*, p. 54.

they lend are the real resources they could acquire if, instead of holding money, they spent it. When the expansion or contraction of bank liabilities proceeds in such a way as to be at all times in agreement with changing demands for inside money, the quantity of real capital funds supplied to borrowers by the banks is equal to the quantity voluntarily offered to the banks by the public . . . Thus a direct connection exists between the conditions for equilibrium in the market for balances of inside money and those for equilibrium in the market for loanable funds. An increase in the demand for money warrants an increase in bank loans and investment. A decrease in the demand for money warrants a reduction in bank loans and investments.²⁹

Therefore, fractional reserve banking avoids excess demand and supply of money because the issues of the banks are virtually irrelevant. It is only their customers who choose the appropriate money balances and thus the total quantity of money in use. Balances are held in consideration of the purchasing power of money, that is, the money prices prevailing on the market. "People who find themselves holding excess notes or deposits will get rid of them largely by depositing them in checking or savings accounts at their own bank, or by spending them away to persons who will deposit them."³⁰ Now, say the free bankers, money prices are exclusively determined by the value of outside money, for example the value of gold. *Money substitutes play no role in the formation of money prices.* The supply of bank money has no influence on the purchasing power of money. Money substitutes must necessarily have the same value as money itself because they are convertible into money. In the eyes of the free bankers, restrictions on the issues of banks would in no manner prevent changes of the price level. This is because the latter exclusively depends on the industrial demand for gold. The following gives a sample of formulations of this anchor theory³¹:

The public's demand to hold the demand liabilities (notes or demand deposits) of any particular bank is a definitely limited magnitude in nominal as well as real terms given that the purchasing

²⁹Ibid., p. 55.

³⁰White, *Competition and Currency*, p. 158.

³¹See also Horwitz, *Monetary Evolution, Free Banking, and Economic Order*, p. 120f; Dowd, *Laissez-faire Banking*, p. 65f.

power of notes and demand deposits is fixed by their redeemability for specie.³²

In the limit, with clearinghouse reserves of base money economized to zero and hand-to-hand currency entirely bank-issued, so that neither the banking-system nor the public holds any base money, the purchasing power of base money would depend entirely on nonmonetary demand for the substance comprising base money. Under a commodity standard the value of the unit of account (a standard unit of the base money commodity) would still be determinate. Under a fiat standard, the value of the unit of account would go to zero (because there is no nonmonetary demand for fiat money), placing the system's viability in doubt.³³

. . . a modern competitive theory of money distinguishes between currency [taken as synonymous with high-powered money] and bank money. The stock of currency at any moment is fixed. That fixed stock of currency together with the demand for currency determines its value. Being convertible into currency, bank money or deposits must have the same value as currency. And given a price level determined by the supply of and the demand for currency, the banking system, without affecting the price level, supplies whatever quantity of deposits the public wants to hold.³⁴

This is sheer fallacy. Money prices on the market are the prices paid in form of money *and* in form of money substitutes. It is this *total* sum (the money supply in the larger sense³⁵) which determines the height of money prices. Yet, fractional reserve banking means that there are fiduciary issues of money substitutes. Then the money supply in the larger sense must be bigger than the money supply in the narrower sense (of money itself). In this case, money prices must be higher than the prices that could be formed with the use of money only.

Suppose I get an additional fiduciary banknote of one ounce of silver sterling from my banker. This banknote permits me to

³²White, *Competition and Currency*, p. 158.

³³Selgin and White, "How Would the Invisible Hand Handle Money?": 1724–5.

³⁴Glasner, *Free Banking and Monetary Reform*, p. 174f.

³⁵This means the quantity of money (money in the narrower sense) plus the quantity of fiduciary money substitutes. The latter always equals zero in a 100 percent reserve system. The money supply in the larger sense, then, always equals the money supply in the narrower sense. Only in a system of fractional reserves can these two aggregates deviate from one another.

satisfy wants that hitherto were not sufficiently important to be considered (they were submarginal). If I pay for a meal in a restaurant with this banknote then, without any doubt, I have affected market prices. In fact, by my very purchase I have formed market prices. These prices would have never come into being without the additional issue of a banknote. Selling the meal to other persons would have required a price reduction to attract submarginal consumers. Thus, without the issue of the additional banknote, the money price of a meal would necessarily have been lower. True, the free bankers might say, but if you only *hold* your money, then no new prices are formed on the market. You then have exercised no influence on market prices. But money is always demanded to be spent. Even if an additional fiduciary money substitute is spent only one time it already has raised money prices on the market.

It is the principal shortcoming of the free bankers not to understand the principles of money-price formation.³⁶ They believe that changes in the purchasing power of money are a matter of the long run.³⁷ This is an error. Their entire conception of *how* those changes come about is futile. On grounds of their doctrine, one cannot even conceive of how changes in the purchasing power of money are *ever* brought about. However, the formation of market prices is definitely not a matter of the long run. Money prices are formed by the use of the supply of money in the larger sense. The larger this supply, the higher are the money prices. It is therefore impossible that relative money prices not be distorted or affected by a change in supply of fiduciary issues. *Each* modification of the supply of money in the larger sense affects money prices *with no delay of time*. Once this is conceded, the anchor theory collapses. The decisive influence that money has on its substitutes

³⁶Consequently, it is not surprising that some advocate the absurd idea that the crisis of the thirties had been the outcome of heavy variations of the value of gold (Glasner, *Free Banking and Money Reform*, p. 222ff). For a critique see the articles by Wiegand, Kemmerer, and North in *Gold Is Money*, Hans Sennholz, ed. (Westport, Conn.: Greenwood Press, 1975). Cause and effect are confused. The value of gold changed heavily because of big variations of the quantity of its substitutes. The same confusion prevails about the variations of the gold price of the 1980s. Gold went up because many market participants expected it to soon become money again. It went down when it became obvious that these expectations were premature. This was partly due to the views of experts who considered it as "a commodity whose purchasing power is subject to violent and erratic fluctuation" (White, *Competition and Currency*, p. 131).

³⁷^a[I]t takes time for changes in spending to influence prices in a general way" Selgin, *The Theory of Free Banking*, pp. 53f.

is by its quantity. The quantity of money determines the quantity of money substitutes that can be issued. This money supply in the larger sense, then, enters into the formation of money prices.

It is characteristic of the entire free-banking program to confuse this issue. They adhere to some mythical idea of price formation through convertibility.³⁸ And they tend to consider quantitative limitations on action as accidents to which the attainment of monetary equilibrium is unfortunately exposed. However, with the myth of the anchor falls the myth of monetary equilibrium and its complements, excess demand and excess supply of money. It is untenable that "short-run corrections in the real money supply require changes in the nominal quantity of money."³⁹ A change of the (nominal) supply of money can never be warranted "because it maintains monetary equilibrium."⁴⁰

One cannot avoid this conclusion, as Stephen Horwitz attempts to do, by merely redefining terms. Horwitz defines a neutral money as not distorting "the determination of relative prices when there are changes in its supply."⁴¹ It would be as meaningful to define the perfect human being as "someone whose mind is not limited by the category of causality."

Definitions are necessary. What is at stake, however, is not our capacity to *invent* definitions but whether the definition in question is useful or not. No definition can be useful that contradicts itself. Whatever names we choose to describe it, a "money that will not distort the determination of relative prices when there are changes in its supply" is a contradiction. Calling this impossibility *neutral money* means nothing else than that we give a name to something that we cannot even conceive of. Discussing the effects of neutral money is therefore as meaningful as the dissemination of accountancy methods in a socialist commonwealth.

It is frequently objected that the relevant quantity of money is indeterminate. From this, it is inferred that the formation of money prices cannot rely as heavily on the money supply as it has been pointed out above. What does this objection amount to? It amounts to saying that existing stocks are indeterminate. Of

³⁸Convertibility taken fallaciously in its larger, meaningless sense. See the section entitled "Money and Substitutes for Money" in a previous section in this article.

³⁹Selgin, *The Theory of Free Banking*, p. 54.

⁴⁰*Ibid.*

⁴¹Horwitz, *Monetary Evolution, Free Banking, and Economic Order*, p. 134.

course this is not true. The supply of a present good is always limited even if there is no one able to say exactly how much of this good exists. Otherwise it would not be a good. Thus, the stock of a medium of exchange is never indeterminate in any relevant manner. Money and its substitutes are no exception. Confusion about the money supply in the larger sense stems from conceptual confusion.⁴²

*Does Fractional Reserve
Banking Favor Investment?*

The spurious doctrine of the equivalence between money and real goods is not only used as a critique of 100 percent reserve banking. It also underlies attempts to prove the expediency of fractional reserves. Because the ultimate end of indirect exchange is always to buy some non-monetary goods, the use of money cannot have any value independent from the value of the latter.

Thus, say the advocates of fractional reserve banking, money is an entitlement to real goods. It *represents* the real funds for which it is intended to be exchanged. But, unfortunately, there need not always be equivalence of the amount of the loanable funds and the money in circulation. The latter may prove insufficient to buy all real savings. *Distortions* would be inevitable when the *real loanable funds* could not be borrowed because there is no corresponding circulating money to buy them. This is where fractional reserve banks step in. In the form of money substitutes they create the corresponding money that otherwise would be lacking. According to Horwitz:

Savers supply real loanable funds based on their endowments and intertemporal preferences. Banks serve as intermediaries to redirect savings to investors via money creation. Depositors give banks custody of their funds, and banks create loans based on these deposits. The creation (supply) of money corresponds to a supply of funds for investment use by firms.⁴³

This is the essence of the free bankers' creed. In their eyes, only part of the whole money supply is relevant for the market

⁴²Unfortunately, such conceptual confusion prevails also in one of the most brilliant expositions of the problems of fractional reserve banking, see, F. A. Hayek, *Monetary Nationalism and International Stability* (London: Longmans, Green, 1937).

⁴³Horwitz, *Monetary Evolution, Free Banking, and Economic Order*, p. 135.

exchanges. Only the part in circulation constitutes a demand for real goods and services. The other part is money held—the proper demand for money. The owners of money held are lenders: “what they lend are the real resources they could acquire if, instead of holding money, they spent it.”⁴⁴ Because the proportions between money in circulation and money held can change there can be a difference between savings and investment. Suppose someone increases his money balance. Holding more money substitutes, he renounces his share of the goods to which the money entitles him. Now the money he holds no longer *circulates* on the market. He saves but nobody invests. What is more important, nobody *can* invest because the necessary medium of exchange has been withdrawn from the market. Accumulation of unsold goods inventories would be the inevitable consequence were it not for the beneficial intervention of fractional reserve banks. They create new money in circulation that will buy the idle goods inventories. Savings and investment are again in accord with each other.

It is not necessary to point out all the fallacies of the equivalence idea.⁴⁵ We only have to examine its basic tenet regarding investment. The free bankers think that there can be a difference between savings and investment. Yet there is no such difference. Savings and investment are always identical. They are merely two aspects of the same action, just as buying and selling are two aspects of the same market exchange. One cannot save without investing, nor is it possible to invest without saving at the same time. Thus, suppose that Jones sells a car against 50 ounces of gold that he intends to hold until his retirement age. Jones has invested in gold. Yet this means nothing else than that his savings are in gold, too. It is immaterial whether Jones keeps his gold in some worn socks or with his banker or someone else. No additional action of any bank is required to make savings and investment equal.

Now, suppose that Jones keeps his gold with a bank on a demand deposit. His banker thinks—because he has been instructed by some clever free banker—that in lending out these idle funds through the issue of a money substitute, he finances a corresponding investment. He gives two ounces to Smith who, in

⁴⁴Selgin, *The Theory of Free Banking*, p. 55.

⁴⁵Note that its application in the context of savings-investment is incompatible with its application to justify the “anchor theory.” However, as both are fallacious we do not have to dwell on inconsistencies between the tenets of the free bankers.

turn, buys a washing machine. By giving this idle money to Smith, does Jones create Smith's washing machine? Does he create gold? Does he create just one present good? Does he create something else than a demand deposit? If the answer is no—and there can be no doubt about that—how is he able to finance an additional investment project, that is, supply it with some present goods? He takes Jones's money to do that. Thus, *he not only takes Jones's savings but also his investment*. Such actions are commonly called robbery.

Our enlightened banker has financed Smith's investment project by robbing Jones. He has not achieved an economic miracle, at least no miracle that no other robber would be capable of. Of course, in our enlightened age, neither Smith nor Jones are aware of the nature of the blessings of fractional reserve banking. Smith eats the cake of Jones and of the other money owners while the latter think that they still have it. For it is not true that by "holding a bank liability, either deposits or currency under free banking, the possessor refrains from redeeming it for outside money."⁴⁶ Holders of demand liabilities are definitely *not* "granters of credit just as are holders of time liabilities."⁴⁷ The possessor believes that he can have both, benefit from the use of a money substitute *and* redemption whenever he wants. This is exactly why money substitutes under fractional reserve banking are so interesting to him. The banker (and some economists) may believe that there is just "a difference of degree and not a difference of substance" between credits given on a base of demand deposits and credits on a base of other credits. But there can be no doubt that not only *is* there a difference of substance but that this difference constitutes, in Murray N. Rothbard's terms, "the nub of the problem" of fractional reserve banking:

a claim—and banknotes or deposits are claims to money—does not involve the creditor's relinquishing any of the present good. On the contrary, the noteholder or depositor still retains his money (the present good) because he has a claim to it, a warehouse receipt, which he can redeem at any time he desires. This is the nub of the problem, and this is why fractional reserve banking creates new money while other credit agencies do not—for warehouse receipts or claims to

⁴⁶Horwitz, *Monetary Evolution, Free Banking, and Economic Order*, p. 135.

⁴⁷Selgin, *The Theory of Free Banking*, p. 62.

money function on the market as equivalent to standard money itself.⁴⁸

If issuing fiduciary money increased “the supply of loanable funds and spurs further economic growth”⁴⁹ there would be, to be sure, no possibility to dispute some beneficial effects of fractional reserve banking. Then the need to redeem money substitutes must appear as an atavistic obstacle for banking. And so would the limitation of money itself. But if one can really imagine “politicians leveraging the Fed into generating short-term output increases to pump up the economy at election time,”⁵⁰ why do we not encourage our politicians to do that *all* the time? Why do we abstain from continually enforcing “temporary deviations of real output from its natural rate?”⁵¹

The answer is: because it is, *even in the short run*, impossible to generate output increases by printing money. Production capacities for future *and* present goods are always limited. If I convert my existing production facilities to the production of more present quantities, then quantities produced in the future will be reduced. If this were my intention then I would successfully increase output. I would err, by contrast, if I believed that I could have more quantities today without paying in the form of less quantities tomorrow. I cannot feel richer having many goods today when I know that I shall starve tomorrow. When I am convinced that it will rain tomorrow I will repair the roof of my house. I do not think a second of taking too long a sunbath to complete the repair today. The additional hour of sunbathing is, in any practically relevant sense, not more than the repair of the roof of my house. Forcing me to behave in another way, namely, to take a longer sunbath today, can in no conceivable manner be more valuable to me.

In quite the same way, it is impossible to provide more loanable funds through fractional reserve banking. Gold held in demand deposits must be considered as savings. This, however, does not mean that its holders renounce their disposition of it. Fractional reserve banks may be necessary—as is central banking,

⁴⁸Murray Rothbard, “The Case for a 100 Percent Gold Dollar,” in *In Search of A Monetary Constitution*, Leland B. Yeager, ed. (Cambridge: Harvard University Press, 1962), pp. 115–6. Reprinted in book form by the Ludwig von Mises Institute, Auburn, Alabama in 1991.

⁴⁹Horwitz, *Monetary Evolution, Free Banking, and Economic Order*, p. 115.

⁵⁰*Ibid.*, p. 131.

⁵¹Selgin and White, “How Would the Invisible Hand Handle Money?”: 1725 n.

too—because it “would make certain lending opportunities profitable that would not otherwise be worthwhile.”⁵² So is a robber necessary to make loans to those who would not otherwise get them. No bank can procure more loanable funds than the public is willing to place at its disposal. The only thing they can do is to *deceive* their customers about the quantities of factors of production that are available.

Printing banknotes and creating new deposit accounts is not the cause of cars assembled, bridges constructed and children educated. Everyone inclined to ignore this will sooner or later be told better by the course of events. A bank always operates as an intermediary of *already existing* funds. It does not create them. Issuing additional quantities of fiduciary banknotes and demand deposits does not increase the quantity of the goods that can be bought with the new fiduciary money. Hence, inflation cannot represent an increase in output. Only if the receivers of higher nominal incomes *believe* that they can have both more goods today *and* more goods tomorrow is the increased quantity of present goods more in their eyes.

Yet this is a blatant error that only becomes apparent at a later stage of the inflation-induced evolution. Those who believed in the blessings of inflation or who ignored the latter altogether will find that the longer sunbaths of yesterday have to be paid for by a wet dining room today. Inflation-created output increases are a contradiction in terms. Not only do they fail to encourage investment, they positively impede it because they cannot but lead to error, that is, to the destruction of investment. There is no difference between fractional reserve banking and government intervention in financial markets. Both “*divert* savings from more to less productive channels.”⁵³

The free bankers are inspired by a spurious problem. It is therefore that their doctrines are as unsatisfying as those of their predecessors. During almost the whole of our century, economists were in search of the causes and consequences of deviations

⁵²Dowd, *Laissez-faire Banking*, p. 48. This is precisely the argument of the central bankers. Goodhart, for example, claims that central banks are necessary “to support the residual, risky, ‘true’, banking institutions, which were undertaking the necessary function of making loans to borrowers who could not otherwise sell their own equity and debt in extant financial markets” (“Are Central Banks Necessary?,” *Unregulated Banking: Chaos or Order?*, Forrest Capie and Geoffrey Wood, eds. (London and New York: St. Martins Press, 1989), p. 18.

⁵³Murray N. Rothbard, *Power and Market*, 2nd ed. (Kansas City: Sheed Andrews and McMeel, 1977), p. 186.

between savings and investment. Nearly all of them overlooked the disposition issue. (Probably they tried to avoid it because it would have led them too near to the concept of ownership; which was deemed unscientific.) So they tried to explain the recurrent crises of capitalism with the wrong tool. Business cycles are a matter of systematic error. Yet, this error refers to the disposition of goods, not to differences between savings and investment. Their unawareness of the disposition issue leads the free bankers to misconceive the argument of Rothbard in support of 100 percent reserve banking. His claim that fractional reserve banking is fraudulent⁵⁴ is in their eyes "more jurisprudential than economic."⁵⁵ They are certainly right that "nothing in a free banking system prevents an individual who desires 100 percent reserve banking from explicitly contracting for it."⁵⁶ Yet nothing in the world prevents people from being foolish. Rothbard's view that banknotes are the legal equivalent of warehouse receipts is *not* "based on what he thinks legal practice *ought* to be."⁵⁷ Rather it is the other way round. Legal practice ought to acknowledge that banknotes *are* substitutes for money and that it is impossible that two persons dispose of the same good at the same time.

*Does Fractional Reserve Banking Convey
a Superior Kind of Knowledge?*

The fundamental economic fallacy of all brands of socialism is the idea that money is not needed for the calculated planning of action. Unfortunately, there is a corresponding fallacy of just the opposite nature, namely, that the use of money provides *something more* than the indispensable instrument of the calculation of action. This conviction is manifest in the naive attempt to create goods by an increase of the quantity of money. It is also apparent in the attempts to attach a special dignity to money because it allegedly conveys a superior kind of knowledge. Recently the conviction that monetary exchange is a social communication process has found an advocate among the free bankers:

Both language and money are ways of extending our perceptual apparatuses beyond the immediate; the difference lies in to what each allows us access. The advantage of a monetarily extended

⁵⁴Rothbard, "The Case for a 100 Percent Gold Dollar," pp. 114f.

⁵⁵White, *Competition and Currency*, p. 156.

⁵⁶Ibid., p. 157.

⁵⁷Ibid., p. 156.

language over language alone (and why the modern socioeconomic order is equally dependent on money, as it is on language, for its emergence and evolution) is that money allows us to utilize not only the articulate knowledge of others but, more important, their knowledge that cannot be put into language.

He then theorizes what kind of information money does convey:

language and money . . . *constitute* the way in which we express [mental] constructs and preferences. Just as we cannot help but think in terms of the words that language provides us, we cannot help but act in the market in terms of the money prices of what we want to exchange. As difficult as it is to communicate thoughts outside of language, so it is difficult to express market-relevant wants outside of monetary exchange.⁵⁸

There is no doubt that money prices constitute an expression of our preferences. However, this is not the point. The point is that they are but *one* expression of preferences and that the latter are revealed in *any* prices, not only in money prices. Yet, if our preferences are revealed by all market prices then it is impossible to claim a particular ability of money to convey them.

But there are still other, more general flaws in the superior-knowledge theory of money: Money is scarce, language is not. The use of money *implies* social cooperation, the use of language is a unilateral act that does not imply cooperation. The *success* of the use of money is based on a fundamental disagreement about the meaning (more narrowly: the value) of money. As with every market exchange, it presupposes only the knowledge *that* the intended act is profitable (more useful than any other action), not *why* it is so. Market exchange rates convey no knowledge apart from the valuations which made cooperation possible. The use of money permits diverging interpretations of the underlying objective conditions of action. This is of no importance for the success of a market exchange. *Every* use of money, by its mere existence, proves that cooperation is possible even if one partner in the exchange is fundamentally erring. By contrast, the success of the use of language is based on a sufficiently similar interpretation of special objects (words and other symbols). Without an agreement upon their meaning *no* success would be conceivable. The use of a language is impossible without the tacit conviction that the

⁵⁸Horwitz, *Monetary Evolution, Free Banking, and Economic Order*, p. 97.

objects of the discourse are perceived (interpreted) in the same way.

The success of trade and money and its importance for humanity is based on the fact that they do not presuppose any agreement between cooperating individuals about the interpretation of their environment. Money and trade rely upon the extreme opposite foundation, that is, diverging attitudes toward the value of objects. Therefore it is not true that "joint production processes require the communicative agreement that money permits."⁵⁹ The division of labor is certainly *facilitated* by language. Yet, language is but a tool to reduce the uncertainty linked to the interpretation of the intention of others; insofar, it resembles not only money but *all* means of action. It is to this wide analogy that Simmel refers in his *Philosophy of Money*. Such an analogy has limits:

The point of departure for the analogy between money and language is to recognize that both mediate social processes; money is the "medium of exchange" for Menger and many others; language is the "medium of experience" for Gadamer and others in the Continental tradition . . . Language and money do not reveal some preexisting mental constructs or preferences, rather they constitute the way in which we express those constructs and preferences.⁶⁰

Simmel's authority, therefore, cannot be claimed in support of the idea that money is—as language—a means of communication.

The difference between 100 percent and fractional reserve banking is of course one of error and information. Yet fractional reserve banking is far from being superior in this regard. Rather the opposite. Under 100 percent reserve banking the factor use linked to the employment of money substitutes shows itself in the costs incurred by the bank customers. Under fractional reserve banking just the opposite holds true each time additional quantities of fiduciary money are issued. The bank customer receiving

⁵⁹Ibid., p. 100.

⁶⁰It seems to be the intention of discussing Simmel's work at length to prepare the ground for a communication theory of money (ibid., pp. 91ff). The same reproach must be made for citing Mises's ideas on the importance of language from his *Nation, State, and Economy* (New York: New York University Press, 1983). Indeed, these belong to the few ideas Mises considerably revised later on (*Omnipotent Government* [New Haven, Conn.: Yale University Press, 1944]). It is impossible to claim his authority in support of the tenet that "ideas do not exist extralinguistically" (Horwitz, *Monetary Evolution, Free Banking, and Economic Order*, p. 186).

an additionally issued banknote or demand deposit at (necessarily) too low a price believes this price to be the costs of the credit.

In fact, this is not the case; the rest of the costs must be paid by the other money owners in form of lower purchasing power. The customer receiving interest payments for his money deposited in a demand balance believes him receives a free lunch. In fact, they do not do so for their deposit with a bank makes fiduciary issues possible and thus leads to a decrease of the purchasing power of their money. Additionally, they are erring about the quantity of money they can dispose of. However, only in times of liquidity crises do those errors on a wide scale become obvious.⁶¹ Thus it is precisely under a regime of fractional reserves that the market participants are systematically *misinformed* about the quantities of goods they can dispose of. It is also unlikely that, under fractional reserve banking, “reserve holdings would indeed fluctuate to reflect the trust that the public holds in a bank’s liabilities and the confidence the bank has in its assets,” as Horwitz believes.⁶² If this interpretation was common in the market then even bad banks—and *especially* bad banks—would do their utmost to operate on a low reserve ratio.

The breakdown of any system of fractional reserves represents only the cluster of failure that was already implied in the cluster of erroneous assumptions concerning the quantity of disposable goods. Insofar as there are striking parallels to the issue of gold versus fiat money, the latter has traditionally been defended with reference to the smaller resource consumption that it would allegedly imply. Yet, at the end of this century, marked by fiat money regimes all over the world, even the most ardent of its champions admit that this was an illusion.⁶³

The Necessary Failure of Fractional Reserve Banking

The Two Sources of Business Failure Implied in Fractional Reserve Banking

The free bankers think that fractional reserve banking can, in principle, last forever. They believe that it does not bear in itself the source of its destruction. They are convinced that its pure existence does not imply its decline. They are wrong on each one of these contentions.

⁶¹See Machlup, *Börsenkredit, Industriekredit und Kapitalbildung*, pp. 143ff.

⁶²Horwitz, *Monetary Evolution, Free Banking, and Economic Order*, p. 146 n. 46.

⁶³Milton Friedman, “The Resource Cost of Irredeemable Paper Money,” *Journal of Political Economy* (1986).

Some critics of fractional reserve banking think that the root of the free bankers' fallacies is that they maintain that the holding of money constitutes savings. Yet, as it has already been stated above: one cannot save without investing, nor is it possible to invest without saving at the same time. This refers to *all* goods and, thus, to money. The terms savings and investment (or better: savings-investment) refer to *all* actions. From the point of view of the acting person each means which he disposes of—even for the shortest delay of time—is savings-investment. Like the category of means-ends, it is a categorical feature of action. The machines owned by a great industrialist are as much his savings-investment as the coffee cup that I own as a part of my savings-investment. So are cars, refrigerators, dentists' equipment, computers, and the fresh pizza served in a restaurant. And so is the money that one owns, too.⁶⁴

However, these considerations are only preliminary to what economics is all about, that is, the employment of *limited* means or goods. In a situation of unlimited means there could be no question of the success of action. All actions would be successful because of the abundance of means. Action could not be as we know it. Just the contrary is true for goods. Only a limited number, representing a limited range of actions requiring their

⁶⁴Here my opinion deviates from that of Rothbard. He says: "A man may allocate his money to consumption, investment, or addition to his cash balance." (*Man, Economy, and State*, p. 678, see also pp. 179f), thus suggesting that holding a cash balance is something different from savings-investment. Hans-Hermann Hoppe has given another expression to this view in claiming that time-preference and the utility of money are "two distinct and praxeologically unrelated factors" (*The Economics and Ethics of Private Property* [Boston: Kluwer, 1993], p. 119). To be sure, there is no causal connection between the demand for money and the interest rate. Increasing the quantity of money cannot reduce the interest rate because money's *real* value, its purchasing power, would be reduced accordingly. Yet this is no reason to overlook the unity in all *acts*, viz., in all valuation. Value is the preference accorded to an effect, and at least in the realm of action this can mean nothing but that the preferred effect should be achieved *before* alternative but less urgent effects. As action—and all other means—are always employed in the pursuit of some ends or effects acting man necessarily has to value (i.e., select) his means according to the urgency of the ends they are supposed to achieve. Thus, time-valuation is present in all actions. Actions with money can be no exception.

However, it should be noted that it is the holding of *money* which constitutes savings-investment. The holding of money substitutes, on the other hand, does not constitute savings-investment but *claims* on savings-investment in the form of money. It *cannot* give disposition of more than the existing stock of money—even if the owners of fiduciary money substitutes *believe* the contrary to be the case. See Böhm-Bawerk, "Rechte und Verhältnisse vom gueterwirtschaftlichen Standpunkt," in *Gesammelte Schriften* (Frankfurt/M.: Sauer & Auvermann, 1968).

use can be successfully executed. This number or range is larger when more goods can be employed, and it is smaller when fewer goods can be employed. Yet at each moment it is limited. The main problem of acting man consists of the identification or discovery of the most important actions which—under the prevailing limitation of goods—can successfully be carried out. *This problem can only find a solution if and insofar as acting man correctly identifies how many goods are at his disposal.* He must fail if he errs in his appreciation of the amount of goods he can dispose of. If an institutional arrangement implies that the acting persons under its influence err systematically, the arrangement itself can be said to lead to necessary failure. This is exactly the character of fractional reserve banking.

Errors are regrettable but there are no known means to avoid them. Error in business consists of a false appreciation of the future values of consumers. It will occur at all times and in all places, with or without 100 percent reserves in banking. In comparison to the totality of all actions, however, error is but a minor phenomenon. Given sufficient time, man learns how to deal successfully with all objects, be they means or obstacles to his ends. One hundred percent reserve banking is no object that implies particular difficulties for action. Errors, then, cannot be a characteristic feature of its use. It is quite another case with fractional reserve banking. Fiduciary issue of money substitutes as such is, of course, not the root of business error. There is no link whatever between the coverage of money substitutes and the correctness of anticipation. But in two respects it is always linked with error.

The first respect is that a situation in which reserves are fractional *can only be brought about* by the issue of fiduciary money, that is additional and therefore uncovered money substitutes. The important feature of this bank-created inflation is that it must lower the interest rates charged by the banks. Without lowering the interest rates, they would simply be unable to lend the additional money substitutes. Considering the lower interest rates, more projects are calculated to be profitable and launched. Yet, because the production capacities are limited, this must lead to a “cluster of business error,”⁶⁵ that is, to approximately synchronous failures of many market participants.

Dealing with a fractional reserve banking system, market participants are permanently misled. In their calculations more

⁶⁵Rothbard, *America's Great Depression*, p. 16.

projects appear to be profitable than can be successfully finished. The nature of fractional reserve banking is to cause this kind of failure on a wide scale. One cannot reproach the free bankers because they do not consider the lessons of Austrian business cycle theory. At least some of them do know that the issue of additional quantities of money substitutes leads market participants to make systematic errors. Yet general errors of market participants do not stem from a confusion "between nominal and relative price changes."⁶⁶ All prices are nominal. Without a denomination in some unit there would be no means with which to compare them.

The second aspect is contagion. Even the free bankers do not deny that under fractional reserves, the failure of one bank is likely "to trigger systemwide runs, implying large-scale demands to redeem banknotes and deposits for base money" leading to "widespread bank failures, undermining the payments system."⁶⁷ Nevertheless they do not believe this to be a devastating critique of their case. They argue that systemic crises in the past have not been a great threat in banking systems. In their eyes it was rather legal restrictions that played a crucial role. They believe that the evolution of an unhampered market would lead to institutions capable of avoiding runs and panics. Let us examine these arguments in turn.

Legal Restrictions and the Exogenous Causes of Bank Failure

Runs on the banking system, it is said, "were precipitated by events exogenous to the banking systems."⁶⁸ Now, what is an *exogenous event*? Imagine a blind person walking without orientation on the pavement. What if he falls in a hole and breaks his neck? One could blithely argue that the reason for his accident was

⁶⁶Selgin and White, "How Would the Invisible Hand Handle Money?": 1725.

⁶⁷Ibid.

⁶⁸Hasan Ifedhar and Gerald P. Dwyer Jr., "Bank Runs in the Free Banking Period," *Journal of Money, Credit, and Banking* 26 (1994): 284. Or, in the terms of A. J. Rolnick and W. E. Weber: "free bank failures were not caused by individuals establishing free banks with the same intention of having them fail. Rather, free banks failed when economic times turned bad and the value of their portfolio declined. Thus, the problems of banks during this period do not appear to have been different from those encountered by banks in other periods or by other types of industries" ("The Causes of Free Bank Failures," *Journal of Monetary Economics* 14 (1984): p. 290. See also Glasner, *Free Banking and Monetary Reform*, p. 203; Dowd, *Laissez-faire Banking*, pp. 218f; Horwitz, *Monetary Evolution, Free Banking, and Economic Order*, p. 152ff).

exogenous to his being blind. He did not break his neck because he was blind but because some unpredictable circumstances *from outside* disturbed his otherwise brilliant fate.

The futility of this reasoning is obvious. The concrete object that hurts the blind is as immaterial for the issue as the concrete reasons that lead too many market participants to redeem their money substitutes at the same time. It is also immaterial whether the concrete causes for failure are defined as exogenous to the activity in question. The only relevant aspect in this context is whether the activity in question implied already certain problems or not. In the above cases there can be no doubt about this point. To be blind *means* to be exposed to the increased danger of collision. To hold fractional reserves *means* to be exposed to the danger of having to redeem more than one is able to. The free bankers think they have refuted the reality of contagion crises. Yet, they have merely played with words. The contagion or domino effect is not refuted if one *defines* it conveniently. There is no use in building up a straw man called contagion crises and meaning a sudden breakdown of confidence in the banking system that comes out of heaven. There is no such thing as contagion in this sense. It is therefore not very surprising that such contagion never occurred in the past and that it will never be easy to find in practice.

At the bottom of the issue is the relationship between the psychology of the actor and the success of his actions. For economic analysis, the importance of a belief is not its mere existence but the conditions under which it leads to successful action, viz., under which it is right. Crises of confidence and bank runs can be interpreted in two ways. Either one has to suppose that the prevailing conditions justify them, that is, render them successful. Considering fractional reserve banking this is undoubtedly the case. Timely redemption always proves to be successful because it is impossible to satisfy all redemption demands. Or one has to suppose that the beliefs of the market participants are completely erroneous. Why, then, do they err all at the same time and in the same way? Are they guided to similar behavior by a somewhat mysterious herd instinct?

To these questions, the free bankers have provided no answers. To be sure, everybody necessarily acts according to what he *believes* is right. But it is quite a different question whether the convictions of the actor *are* right, that they too lead to successful action. Does the existence of an individual belief, or

confidence as such, imply that it be justified? Is a bank illiquid *because* the depositors believe it to be so? Implicitly, the free bankers answer these questions in the positive. For if contagion crises are “crises of confidence” and contagion effects are “confidence externalities,”⁶⁹ one is led to believe that the mere breakdown of confidence in banks leads to a breakdown of payments.

At least under 100 percent reserve banking this is obviously not the case. Here there could be crises of confidence, but there can be no crises of the payments system. This is because the monetary aggregate that is relevant for payments—the money supply in the larger sense, that is, money plus fiduciary issues—could not differ from the supply of money. Its quantity could only vary to the extent that the quantity of money varies. At least in the case of gold this is of no practical importance.

Whether the money at the disposal of the market participants is in the vaults of the banks or under grandmother’s pillow is, under 100 percent reserve banking, of no importance for this aggregate. Hence, contagion as suggested by the definition of some free bankers cannot be relevant for banking crises.

Why Fractional Reserve Banking Must Always Lead to Bank Runs

Under a fractional reserve banking system it is impossible to redeem all money substitutes. If a bank that has issued fiduciary money substitutes is forced to redeem more of its substitutes than money in its vaults, it has but one option to avoid bankruptcy. It has to borrow the money from other banks. The latter, thus, are confronted with the following dilemma: either they do lend the money, thereby depleting their vaults and becoming themselves illiquid, too. Or they refuse to lend the money and the former bank goes bankrupt. Then all of its customers—but *especially those who have not been able to redeem the substitutes in their possession*—will try to get some cash. They will search for money or money substitutes from the remaining banks. The deceived customers of the ruined bank need cash to maintain their daily transactions. They have no money to deposit, but they need money or money substitutes right now. The remaining banks, however, are not able to accept them as customers. Their stocks of money have not been increased. The issue of further fiduciary money would inevitably make them illiquid. But even this refusal to issue additional quantities of fiduciary money

⁶⁹Selgin and White, “How Would the Invisible Hand Handle Money?”: 1726.

cannot avoid their ruin. They are doomed, too. For if those deceived customers of the first bank cannot immediately dispose of cash they go bankrupt and thus cause liquidity problems for their creditors. Now the latter will have to ask the remaining banks for more cash, creating the old problem on a wider scale.

The contagion effect can only be stopped one way. There must be a bank that is able to satisfy all demands of redemption. However, if the failing market participant is big enough, contagion cannot be stopped at all, at least not if money production is as costly as in the case of gold.

One could ask whether it must necessarily come to a situation in which one single business failure proves to be too big to be borne by the banking system. The answer is: the principle of fractional reserve banking brings it about. Each banker can successfully operate on the hypothesis that in the case of a personal liquidity crisis, he can rely on his fellow bankers. It is in their interest to save him to avoid a bank run. Under such circumstances, the permanent expansion of fiduciary issues provides almost riskless profits. These are the objective conditions of fractional reserve banking. Even the less clever among the bankers will discover them after some years of business experience. Even the less alert among them will behave accordingly, that is, try to reduce their reserve ratio as far as possible. This expansion makes an individual failure ever more dangerous because the reserve ratio is further and further reduced.

It is the possibility of this expansion, however, which the free bankers deny. They claim that there are at least two obstacles for a bank willing to expand its fiduciary issue. The first obstacle to their note issues would be limited by the demand to hold them.⁷⁰ Banks are only capable of issuing according to the demand of their customers. All money substitutes that the latter did not really want to hold would quickly be returned to the bank and their redemption be demanded.

Undoubtedly it is true that all money substitutes held by the market participants are really wanted. Neither can it be disputed that each redemption of a money substitute means that its owner does no longer want to hold it. This, however, is completely beside the point. The only question is whether the demand for money—and, thus, for its substitutes—is limited or not. Yet it

⁷⁰Dowd, *The State and the Monetary System*, p. 62. See also the references given above in the section entitled "Does Fractional Reserve Banking Lead to Monetary Equilibrium."

certainly is not. Let us avoid any misunderstandings. Demand in the sense that the free bankers use this term means desire to dispose of money substitutes, it does not refer to "effective demand, to desires made effective by being 'demanded', i.e., by the fact that something else is 'supplied' for it."⁷¹ It is the very intention of the free bankers to put liquidity at the disposal of market participants without forcing them to supply it. Considering the demand for money in this sense one has always to remember that money is a *present* good. It can be used now. No present good is available in a quantity that would satisfy *all* demands. This is precisely why it is a good. Hence, there is always demand for some more money to secure hitherto less important (submarginal) satisfactions. It is correct that under fractional reserve banking "market forces compel banks to issue more money, when, at given prices, more of it is demanded by the public."⁷² But that demand is unlimited.⁷³ It therefore cannot limit the issues of fiduciary money.

However, the free bankers might say that the expansion of fiduciary money substitutes encounters still a second obstacle that will limit it. That is any expansion increases the risk of depletion of the money stock of the bank. True, but how does our banker know how much he can increase without going bankrupt? There certainly *is* some point beyond which his costs increase "faster than revenue, and so expansion beyond that point is unprofitable."⁷⁴ Yet, it is not the *existence* of such a point that is the problem, *it is the ignorance of its exact location*. No banker knows and can know exactly in advance what amounts of issues are still profitable and which prove to be ruinous. There is but one means to find it out: trial and error. This is, to be sure, the foremost principle of all action. But *in all other businesses than fractional reserve banking saving a competitor is no condition of one's own success* because individual failures (and follies) do not systematically lead to the breakdown of the whole industry. Fractional reserve banking is different. The reserves of the bigger banks may suffice to ignore bankruptcies of some minor competitors.

⁷¹Rothbard, *Man, Economy, and State*, p. 677.

⁷²Selgin and White, "How Would the Invisible Hand Handle Money?": 1725.

⁷³It is unlimited without regard to the prevailing money prices on the market. The latter, however, rise every time the quantity of fiduciary money substitutes is enhanced. They must necessarily be higher than they otherwise would have been. The circumstance, too, leads to higher demands for holding money.

⁷⁴White, *Competition and Currency*, p. 25, Cf. also Selgin, *The Theory of Free Banking*, p. 46.

Liquidity problems of big competitors, however, cannot be ignored. Every banker knows this. Every banker knows that it is in the interest of his fellow bankers to save him. Hence, he has every reason to be audacious in the exploration of that point beyond which the expansion of his fiduciary issues is unprofitable. And at least the alert customers of the banks do know this, too. They will always be very alert for news indicating probable bank insolvency. Thus they will quickly redeem their money substitutes to protect themselves. Taking these precautions they behave in no conceivable manner “contrary to the theory that depositors stage runs simply out of fear that others might run.”⁷⁵

It is of no relevance that the market participants have less *confidence* in their business partners, be it banks or others. It is not important *where* the chain of failures sets in—in a bank or in some exogenous institution. All that is needed is that the error be sufficiently big to cause a sufficiently big bank to fail. Then a succession of failures cannot be avoided. Fractional reserve banking is frequently seen as a kind of multiplier of reserves. In fact, it is a multiplier of error. Fractional reserve banking is an iron chain that links the errors of one or a few market participants with the errors of all the others. Under 100 percent reserve banking, too, there may be some banks that engage in lending operations based on maturity transformation. This, however, is no characteristic feature of 100 percent reserve banking. But it is the essence of fractional reserve banking. Here *all* the banks by the nature of their operations are exposed to the risk of having to redeem claims of others without yet being entitled to demand the redemption of their claims.

Liquidity Crises in the Past

Relying on past events can often be helpful to illustrate political and theoretical issues. However, it can provide no evidence. Even if no failure of fractional reserve banking had occurred in the past this would be no proof that sooner or later it will not have this consequence. Therefore, two notes on this subject will have to suffice.

A central problem of the study of history refers to the *evaluation* of events. There are free bankers, for example, who consider three suspensions of payments in about 50 years time to be not much.⁷⁶ From the point of view of an enlightened economist this

⁷⁵Selgin and White, “How Would the Invisible Hand Handle Money?": 1726.

⁷⁶*Ibid.*, p. 1726. They note that *only three* out of six major panics in the National Banking era involved suspensions of payments.

may be true. The point of view is different for someone who lost all or some of his wealth in one of these three panics. He suffers from an act of deception. He is ruined because his banker committed—willingly or not—fraud on him.

The free bankers have accorded much attention to the relative success—absence of any major crises—of the Scottish-fractional-reserve banking system of the first half of the eighteenth century. The critics of fractional reserve banking have pointed to the dependence of the Scottish banks on the financial city of London. The latter in turn depended entirely on the Bank of England. Thus, Scottish free banking was not free at all, but a remote part of the English central banking system.⁷⁷ The free bankers deny this. In their eyes the “Scottish banks did buy and sell assets in the London financial market, but did not hold deposits at the Bank of England nor, it seems, any significant quantity of its notes. Nor did the Bank of England make last-resort loans to the Scottish banks.”⁷⁸

Yet it is immaterial whether the Bank of England was *directly* involved in securing money for the Scottish banking system. In times of trouble the Scottish banks could always rely on credits from London banks. The huge London market could always provide money if sufficient interest was paid. Thus it is because they resorted *indirectly* to issues of the Bank of England that the Scottish banks depended on the latter as well.

Contractual Remedies I:

Option Clauses, Equity Claims, and Monetary Disintegration

The most striking contradiction in the free bankers' program is their grudging confession that it is unpracticable. No free banker disputes that the suspension of payments is the ultimate recourse of fractional reserve banks.⁷⁹ Yet, redeeming its money substitutes is no generous favor that a bank renders to its customers. Redemption cannot be suspended like granting credit. The inability to redeem is what constitutes bankruptcy. In *all* businesses it is the inability to pay money owed that constitutes bankruptcy. The free bankers, by contrast, believe that

⁷⁷See Murray N. Rothbard, “The Myth of Free Banking in Scotland,” *Review of Austrian Economics* 2 (1987): 229–45. See also Charles Goodhart, *The Evolution of Central Banks* (Cambridge, Mass.: MIT Press, 1988), p. 51f.

⁷⁸Selgin and White, “How Would the Invisible Hand Handle Money?": 1732. See also Lawrence White, *Free Banking in Britain: Theory, Experience, and Debate, 1800–1845* (Cambridge: Cambridge University Press, 1984); Dowd, *Laissez-faire Banking*.

⁷⁹Cf., e.g., Selgin, *The Theory of Free Banking*, p. 137; Glasner, *Free Banking and Monetary Reform*, pp. 199ff.

this inability could just be a tiny little liquidity problem. There may be banks, they say, which essentially are solvent. These banks just need some time to provide the liquid funds to pay out their impatient and ill-informed customers.⁸⁰

This argument ignores the fact that time is a good. If we always disposed of just a little bit more time we could be sure to have reached nirvana. With always just a little bit more time one could provide all the money in the world. Unfortunately, every means in the mundane life of the human race is limited. Time, therefore, plays a crucial role for the success of action. In every place outside nirvana one has to pay for the time-saving means called goods. There is no possibility of providing "liquidity to the market only."⁸¹ One cannot pay with liquidity; one can only pay with goods.

Yet who pays for the banks if they are unable to pay for themselves? The free bankers reply that the bank customers might *agree* to pay for the banks. They might accept devices (such as option clauses and the transformation of money substitutes into equity claims) permitting the temporary suspension of payments. Thus the fractional reserve banks could always stay in business without ever violating contracts. It is very doubtful whether these contractual remedies would be contractual legitimations of fractional reserve banking.⁸²

For the sake of the argument let us assume they would. However, they cannot be *remedies* for the shortcomings of fractional reserve banking. They merely permit banks to cure the liquidity problem by the issue of further fiduciary money substitutes, saving the banks at the expense of the other market participants. If this is a remedy then it is a very general one. Applying the same argument one could say that robbers merely solve their liquidity problems. Or imagine an engineer supplying motors that always explode. It is conceivable that he finds buyers for his products even if he warns them. Yet this does not change the fact that his motors *do* explode. If other people are damaged—which in the case of those motors will not occur as inevitably as in the case of fiduciary money—the engineer could argue: It is not only me and my customer who profit from the use of my motors. You profit from it, too, because the prices I charge are

⁸⁰Cf. Dowd, *Laissez-faire Banking*, p. 48.

⁸¹Selgin and White, "How Would the Invisible Hand Handle Money?": 1727.

⁸²For a refutation of this claim of the free bankers see Hans-Hermann Hoppe, "How is Fiat Money Possible?": 70f.

lower than those of my competitors. My motors are worse, to be sure, but the factor use in their production is lower. Hence, everybody profits from my product and from its exchange on the market. Forcing me or my customers to pay indemnities now is tantamount to ruining me. Then nobody will profit anymore. Give us just a little bit of time and let us continue our business. Then we shall be able to pay indemnities to everyone.

Nobody would accept such a proposal. If there is just one person suffering from the effects of the explosion the owner of the motor would have to pay an indemnity and stop using the motor. In the realm of banking another kind of law seems to prevail. If just one market participant does not give his consent to fiduciary issues and uses money instead his rights are violated. Yet, nobody is forced to pay indemnities and nobody is forced to abandon fractional reserve banking.

All alleged remedies for fractional reserve banking have one thing in common: they seem to shift the frontier separating efficient and inefficient enterprises. They seem to retrieve some banks from liquidity crises that could not otherwise be salvaged. They promise the age-old economic would-be miracle of rendering submarginal projects profitable with more money, without more work, productive innovations, and savings-investment. This is, of course, an illusion. The quantities of all goods are always limited. Contractual remedies per se do not create new goods. They can save the banks—but the bill has to be paid by the other market participants. Option clauses, deposit insurance, and the transformation of money substitutes into claims on equity of the banks all imply higher inflation. Yet inflation is not costless. It is tantamount to prescribing higher doses to a drug addict, thus ruining him further. A drug addict, though, inflicts harm only on his property. The contractual remedies recommended by the free bankers harm even those who did not give their consent.

How can one seriously advocate a system without believing in its success? The free bankers do not torture themselves with questions of this kind. In strict accordance with the principle that if reality does not comply to theory then it is a poor reality, they propose to take another attitude to life itself. Selgin and White, for example, suggest that the unconditional demandability of banknotes and some deposit liabilities may be the result of legal restrictions rather than market forces:

Discussions of bank runs and panics ordinarily assume that a bank continues to pay out base money until either all demands

are satisfied or the bank is declared bankrupt. An alternative exists: a bank may suspend payments of base money before such payments render it insolvent and force it into bankruptcy. Although suspension is often regarded as inherently a violation of a bank's contractual obligations to holders of its demandable debt, the unconditional demandability of banknotes and some deposit liabilities may be the result of legal restrictions rather than market forces . . . Under *laissez faire*, bank liabilities might be conditionally demandable only.⁸³

Yet the point is not whether the restrictions of the use of money substitutes are legal but whether they can be removed by an act of legislation. It is devoid of any sense to attempt a definition of legal restrictions covering any conceivable obstacle to any action. No means can be used in the pursuit of opposite ends at the same time.⁸⁴ When I use my shoes to take a walk in Central Park you cannot burn them to heat your kitchen in Montana. Yet this is certainly a restriction of your actions. There is no difference in regard to money. Each use of an ounce of gold must exclude other uses which could be made of it. The legal interdiction to issue more claims to money than money exists merely acknowledges this fact.

Last but not least, no advocate of option clauses seems to be aware that as soon as they are used, a system of different *moneys* is established. The same holds true for all essays to link checking services to equity claims. When money substitutes cease to be claims on money and, though, continue to be used each of them constitutes a different price system. Before, all of them were just expressions of the disposition of money. Thus there was just *one* price system. Now, using them does not mean any more use of money. A general acceptance of such devices would thus lead to monetary disintegration.

Contractual Remedies II: Central Banking and Inflation

The only means to avoid monetary disintegration while preserving the principle of fractional reserve banking is to pool the money reserves. Contrary to the conviction of the free bankers⁸⁵ it is immaterial which form of cooperation this pooling takes. It

⁸³Selgin and White, "How Would the Invisible Hand Handle Money?": 1729.

⁸⁴On this point see Hoppe, *The Economics and Ethics of Private Property*, p. 14.

⁸⁵Cf. Selgin and White, "How Would the Invisible Hand Handle Money?": 1732f.

can take the form of private-branch banking. It can also take the form of a private-central clearing house or of a government-imposed central bank. In each case the effects caused by concentration of money that, before, was dispersed are the same. Pooling permits a shift to large quantities of money to satisfy large but isolated redemption demands. Crises that, before, emerged out of a local liquidity problem can now be prevented. Redemption demands that, before, were critical for the whole monetary system can now be satisfied. Thus, apparently the necessary condition to stop a contagion crisis is now given. Finally, one bank seems to be able to satisfy all redemption demands.

However, one must not overlook that these effects are caused by the *pooling* of money, not by money pools as such. They are merely temporary. Pooling, therefore, cannot avoid bank runs forever. Because there are now greater facilities to provide liquidity the banks will expand their fiduciary credits, thus reducing the reserve ratio again. Only for the time needed for this expansion can the pooled stock of money suffice to help even the biggest banks out of liquidity problems.

Sooner or later, however, the reserve ratio will be reduced to such an extent that the old problem appears on a new scale. Redemption demands that, before, were uncritical now become critical for the whole monetary system. Some banks become big enough to cause, by their failures, crises of the whole system. Hence, the pooling of money stocks does not change the underlying problem of fractional reserve banking. Its main effect is to keep bankrupt banks in business and to make the other market participants pay for it. Not only are the banks able to continue the issue of fiduciary money substitutes, they can even expand it. They grow, not by increasing their services but by expropriating the other market participants.

As no final relief can be brought about by the pooling of money stocks there are but two options for the management of the pooling institution. Either it has to break the redemption promise or it has to look for possibilities to profit from a further concentration of the money stock. This was the problem faced by the central banks during the time of the old (fractional reserve) gold standard. Suspension of payments by the central bank causes principally the same effects as suspension by a single bank. If its money substitutes continue to be used they take the place of the former money. The ensuing monetary disintegration will inevitably reduce the division of labor and permit the central bank to inflate almost at will. This is, of course, the situation we find

today. On the other hand, a further concentration of the money stock must sooner or later lead to a pooling institution on a world scale. Then, at least, there would be no further solution to the persisting liquidity problem of fractional reserve banking than to break the redemption promise once and for all time. While this would have no disintegrating effects on the division of labor it would eliminate all obstacles for inflation.

However, the power linked to a (world) fiat money can only be preserved as long as hyperinflation does not result. Yet hyperinflation is inevitable if the banks are not prevented from ignoring liquidity constraints. There is but one efficient means to assure this: to regulate the free-banking, fractional-reserve, fiat-money system, that is, to impose violent restrictions on this business and especially on the credit volume. Of course, no legislation can prevent the reduction of the reserve ratio. Typically it forbids credit contracts that the rulers pretend to be especially risky. It thus makes banking more bureaucratic, suppresses competition, and, contrary to its intentions, shifts the credits into more risky investments. Thus ever more regulation becomes necessary to suppress its own unintended consequences. The contractual remedies proposed by the free bankers are roads that lead to nowhere. Far from representing solutions they aggravate the problem. They force all other market participants to patronize a destructive system which sooner or later will lead them to hyperinflation or socialism.

The Necessary Failure of Fiat Money

How Gold Becomes Money in an Unhampered Market

The above sections have dealt with the monetary issues of banking. It has been shown that the case for fractional reserve banking is weak. The free bankers' arguments against 100 percent reserves, as well as their arguments for fractional reserves, are wholly untenable. The same holds true for money proper. Here the free bankers display the same inflationist predispositions, viz., their dissatisfaction with gold. Gold is criticized because its supply is not flexible, that is, not as inflationary as its opponents would like it to be. Of course this criticism is spurious on the same grounds as the case against 100 percent reserves.

The quantity of money does not determine the benefits of its use. *All* variations of its supply are harmful. The only qualification to this statement is the increased non-monetary benefits that stem from an increased supply of specie. However, it is not sufficient to prove the

case against gold to be unfounded. One also has to show that the case for other moneys is futile. Can there be a case for fiat money? Is it conceivable that such a system could be successful?⁸⁶

Money is exchanged to be exchanged again in the future. It is bought in exclusive consideration of its future purchasing power. Yet the only successful technique for the estimation of future prices is to base this estimation on present prices, that is, the prices of the immediate past. Today's money prices, therefore, will always rely upon money prices of yesterday. This is the meaning of Ludwig von Mises's regression theorem.⁸⁷ It has vast implications for the theory of money.

Its most important implication for the analysis of the competition between moneys is that it is impossible to introduce new moneys out of thin air. History has featured just one technique for the introduction of new moneys. First, one issues documents representing a claim on money. These documents can become money substitutes if their owners can redeem them at par whenever they want. Yet, their circulation is restrained if they have the character of certificates because in this case a price has to be charged for their use. Once there are *fiduciary* issues, however, money substitutes can crowd money out of circulation. Whenever this happens, the opportunity has come for would-be entrepreneurs to *introduce* a new money. Their method is simple: they break the promise they gave and refuse redemption of the documents they issued. The latter can stay in circulation because there are already prices for them on the market. Yet, such an obvious violation of property rights on a wide scale is only possible if government does not assume its duty to punish that entrepreneur. Past governments have not only spared such persons from prosecution, they have often protected them or were even identical with them. Clearly the necessity of recourse to like procedures for the introduction of a new money represents an important limitation on competition in the realm of money. It is especially this practical aspect that has been completely overlooked by its champions.⁸⁸

⁸⁶For the following see the third part of my *Logik der Währungskonkurrenz*, forthcoming from Frankfurt/M: R. G. Fischer, 1996.

⁸⁷See Ludwig von Mises, *Theorie des Geldes und der Umlaufsmittel*, 2nd ed. (Munich and Leipzig: Duncker & Humblot, 1924), pp. 85ff, also published in English as *The Theory of Money and Credit*, H. E. Batson, trans. (London: Jonathan Cape, 1934); idem, *Human Action*, pp. 408ff.

⁸⁸In his *Denationalization of Money*, 2nd ed. (London: Institute for Economic Affairs, 1978) Hayek simply skips the problem that the Ducates, which he wants

If there is more than one kind of money in use, the regression theorem has to be qualified in an important respect. There is, to be sure, still no possibility of introducing a new money out of nothing. It is still decisive for a market participant who is offered two moneys, A or B, to know at what exchange rate he can sell either of them in the future. The appreciation of this future exchange rate still has to rely upon past prices. But now another determinant of future money prices enters the scene. It is by his very decision to buy money A and not money B, that is, to use A and not B, that a market participant determines the future array of A- and B-prices. If he buys A instead of B he causes a tendency of B-prices to rise and a tendency of A-prices to fall. This means that the exchange rate A to B must rise in which case there would be incentives for him and other market participants to use A and to sell B. This in turn would accelerate this evolution further until B would be driven out of the market and A the only money left in use. In other words, *the simultaneous employment of more than one money implies that each market participant, by his very action, determines the success of this action.* It is his anticipation *per se* that favors its own correctness. This can hardly be said of any other action. In the competition of existing moneys, thus, the progressive character of money-price formation (its orientation to future selling prices) is not only reinforced; it becomes a factor of success of its own. If the competing moneys can be handled with the same ease then this is the only mechanism by which one money can become supreme and drive all others out of employment.⁸⁹

This self-accelerating process cannot be stopped by the fact that the market participants have often opposite views on future selling prices of the moneys in use. Indeed, there may be some who buy A because they expect A to rise and B to be driven out while others buy B because they expect the opposite. Even speculative activities to bring about a rapid fall in one of the moneys can possibly be equilibrated by activities of the same nature but of the

to introduce by a redemption promise, can only become money if this promise is broken. In fact Hayek's Ducates are money substitutes and not money. Otherwise they could never be issued. White holds the same misconception. See also White, *Competition and Currency*, p. 132. For a critique of Hayek's ideas on the introduction of moneys, see Martin Hellwig, "What Do We Know About Currency Competition?," in *Zeitschrift für Wirtschafts Sozialwissenschaften*, 105, pp. 565ff.

⁸⁹The selection of media of exchange of our hitherto non-monetary commodities. See for this mechanism Carl Menger, *Money*, in *Collected Works*, Vol. 4, F. A. Hayek, ed. (1933–36; London: London School of Economics, 1970), esp. chap. 8, sec. 1.

opposite intent. However, once the exchange rate begins to move clearly in one direction it is impossible to prevent the outcome described above—unless there are obstacles hindering the self-fulfilling anticipation of future money prices. Now there are two—and only two—types of such obstacles. The first refers to non-monetary employment in which a money can be used. The second concerns the number of persons who exclusively use either A or B as money.

If a money cannot be used for other purposes than for indirect exchange there is incentive to buy it even at a very low exchange rate. This is obvious in the case of a pure-sign money—as signs do not even have a substance. It is also practically the case with a fiat paper money. One certainly could find *some* employment for mountains of printed paper (burning them for heating purposes, for example). Yet the costs of these actions are likely to outweigh the benefits which could be derived from them. On the other hand, the purchase of gold and silver can never be a complete failure. They are used for many non-monetary purposes—even when their employment as money is suppressed. Gold profits particularly from its physical properties:

platinum, palladium and other precious metals are industrial metals in the possession of dealers and producers, which limits their marketability and deters their use as money. Even silver cannot compete effectively with gold because its current production, relative to its visible supply, is large, exposing its value to sudden changes in quantity. No other metal has such large stock-piles and small current production as gold. No other commodity enjoys as much universal acceptability as gold.⁹⁰

However, one could claim that there still was the second obstacle for the complete abandoning of a fiat money. If there are market participants who exclusively use one money, the exchange rate of the latter can never fall indefinitely. It could always be sold to one of these persons. One could always get a useful commodity in exchange for it. Now, as a matter of fact fiat money is never the only money in use. At least gold and silver are used everywhere and by nearly everyone in the world.⁹¹ As a consequence

⁹⁰Hans F. Sennholz, *Money and Freedom* (Spring Mills, Penn.: Libertarian Press, 1985), p. 67.

⁹¹It is futile to cite the German hyperinflation of 1923 as indicating “that inflation can reach mindboggling proportions before alternative currencies can gain a foothold,” (White, *Competition and Currency*, p. 132). For anyone acquainted with the

gold and silver market prices are omnipresent. Fiat money, therefore, cannot stay in the market if exchange is free. It can never outcompete gold and silver because the latter are also used for many non-monetary purposes. On the other hand, once it is outcompeted by them it can never peacefully come back. It is only preserved because the use of gold as a medium of exchange is systematically suppressed by legal tender laws, regulation of banking and financial markets, and by taxation in fiat money. Therefore, it is wrong to suggest that “neither gold nor inconvertible private currencies will emerge as money under present circumstances.”⁹²

*Transition Toward a Free Money Supply:
The Chimera of Competitive Policies*

Changing the monetary constitution to bring about a free banking system would imply the exclusion of government intervention from money and banking. Yet it is important to pay some attention to the precise meaning of “depoliticizing of money.”⁹³ Abolishing central banks would not lead to a system that was unpolitical in the sense that the banks would not affect the success of other market participants. It would be unpolitical in the sense that it would not be managed by the state, the agency of violent means.⁹⁴

Abolishing central banks would *lead* to a system without government meddling with money. However, the *act* of abolishing central banks would favor some forms of free banking and necessarily prevent other forms. One cannot avoid performing a last measure of monetary policy in abolishing monetary policy altogether.⁹⁵

German mentality of this time it is rather “mindboggling” that even blind trust in authority and heavy penalties could not prevent the use of all sorts of other moneys.

⁹²White, *Competition and Currency*, p. 131.

⁹³Cf. Dowd, *The State and the Monetary System*, p. 185ff; White, *Competition and Currency*, p. 91ff.

⁹⁴For the distinction between economical and political means see Franz Oppenheimer, *The State* (New York: B. W. Heusch, 1914), pp. 24ff.

⁹⁵For the same reason there can be made no vital distinction between rules and discretion as principles of the conduct of monetary affairs. Every rule prescribing *ex ante* how much money has to be issued at what times and in what places and circumstance is discretionary by the very fact that it has to be set up by *someone*. A rule specifying, e.g., different behaviors of central bank officials according to different circumstances cannot even be said to be more “stabilizing” than any pure discretion on their side. It is therefore that rules vs. discretion is a false dichotomy, not because fractional reserve banking has been overlooked as a third alternative (as suggested by Horwitz, cf. *Monetary Evolution, Free Banking, and Economic Order*, p. 125f.

Hence, one cannot avoid answering the crucial question: what money system do I want? The free bankers feel very uncomfortable about this. They are embarrassed by the necessity to choose, that is, to discriminate and they wish to circumvent this problem by permitting competition.⁹⁶ They do not see that one cannot create an amorphous entity called competition and thus remain neutral to the whole issue *Whatever decision one will take, it will necessarily be a decision in favor of something*. Now, money competition will unavoidably lead to the expulsion of all fiat moneys. Even the creation of a world central bank (and thus of one world fiat money) could not prevent it. If this is correct, why not directly choose it? Is it a viable argument that “more than 50 years of being off the gold standard cannot be shrugged off? The past status of gold is not sufficient to guarantee its reestablishment as money?”⁹⁷ Let us disregard this fact that the western world is merely some 20 years off the gold standard. Let us skip for a moment the fact that the use of gold as money is suppressed. The real issue is: what are the alternatives? Can fiat money be said to favor freedom more than gold? Can fiat money persist at all? As long as these questions have to be answered negatively there is just one case for abolishing central banking. This is the case for gold. If there was no gold one would have to invent it. It is correct that a “return to gold without an end to the monopoly of currency issue would at best be half a solution.”⁹⁸ But the same holds true for the inverse relation. Without a return to a 100 percent reserve gold standard, free banking would be far from a full-fledged solution.

A Banking System Which Works and Banking Systems Which Do Not Work

It is bizarre to follow a discussion of “devices for reducing the likelihood that a bank will be unable to provide a full payoff to the last customer in line”⁹⁹ with 100 percent reserve banking hardly mentioned. The free bankers claim that freedom means to place no restrictions “on the terms of contracts made between banks and their customers, beyond the requirement that they

⁹⁶[T]he choice . . . ought not to be foreclosed by anticompetitive policies” (White, *Competition and Currency*, p. 162). Unfortunately this attitude is not limited to the ranks of the free bankers. See also Sennholz who seeks “merely freedom” (*Money and Freedom*, p. 77).

⁹⁷White, *Competition and Currency*, p. 130.

⁹⁸Ibid., p. 135.

⁹⁹Selgin and White, “How Would the Invisible Hand Handle Money?": 1730.

adhere to the standard legal principles governing all business contracts.”¹⁰⁰ This is exactly the point. Yet, it is certainly not such a principle to permit—conscious or unconscious—robbery. The only possible conclusion concerning *legal principles* that justify fractional reserve banking would be that these principles themselves are wrong.

The claims against 100 percent reserve banking are fallacious. So are the alleged advantages of fractional reserves. The principal objection, however, is that neither fractional banking nor fiat money are viable options for action in society. Either they must regularly perish (and each time pull the whole economy with them into disaster) or the payment for the errors they provoke must be coerced by ever increasing state intervention. Hence, the choice at stake is between capitalism and another road to serfdom called fractional reserve banking. One cannot have both.¹⁰¹

¹⁰⁰Ibid., p. 1719.

¹⁰¹For plans to attain free banking on a 100 percent gold standard see Mises, *The Theory of Money and Credit*, pp. 485ff and Rothbard, “Aurophobia: or, Free Banking on What Standard?": 107f.

Portfolio Management of the Free Banks of Illinois: An Examination of Historical Allegations

Salim Rashid and Abdus Samad

The period between 1836 and 1863 constituted an era of free banking in the history of the United States. During this time, withdrawal of regulations by federal and state governments left the banking industry relatively free. Any person or association of persons could open a bank without legislative charter or permission. The opening of a bank was relatively easy. The only restriction was the capital requirement which was limited to not less than \$50,000. The liability provision of the bank was "limited." The shareholders of the bank were liable to the amount of their stocks. During the early stage, banks could be opened in any city, town or village without any population restriction. However, in 1854 and 1857 restrictions of varying numbers were imposed. For example, a bank could not be opened in a locality with a population of less than 1,000. As a result of the relative ease of banking there was a large growth of financial institutions. In Illinois, financial growth was noticed in the operation of 141 banks and their increased volume of money supply and deposit transactions.

The literature on free banking has sharply altered its focus in the last two decades. While the alleged failures of free banking have been closely analyzed and shown to have been greatly exaggerated the link to fractional-reserve banking has been missed in the literature so far. Economopoulos (1988), who pioneered the

Salim Rashid is professor of economics at the University of Illinois at Urbana-Champaign and Abdus Samad is professor of economics at Governor's State University. We are grateful to Andrew Economopoulos and to Charles Kahn for comments; to an anonymous referee for several helpful suggestions; and to the Research Board of the University of Illinois for providing support.

The Review of Austrian Economics Vol. 9, No. 1 (1996): 55-76
ISSN 0889-3047

modern study of Illinois free banking, mentions this issue and we would like to draw out its full implication.

During the period between 1852–1863 a total of 131 banks closed in Illinois. So it was alleged that free banking led to bank failures and “wildcat” banking, i.e., that banks were engaged in issuing liabilities without any cash reserves in their vault or capital and assets to protect these liabilities. Note holders in particular were at risk. In order to avoid payment of the liability holders’ claim, banks were said to have resorted to various unfair practices such as holding fake specie reserve and locating bank offices in an unknown or inaccessible place where they would be difficult to find. Here are a few examples of such views:

Banks of very dubious soundness would be set in remote and inaccessible places where “wildcats” thrived. Banknotes would then be printed, transported to nearby centers, and circulated at par. Since the issuing bank was difficult, often dangerous, to find, redemption of banknotes was, in this manner minimized. (Luckett 1980)

Speculators bought bonds, issued notes to pay for them, and eluded their debtors by taking to the woods among the wildcats. Notes were issued by bankers with no known business. Their (banks) cash were sometimes of nails and broken glasses with a layer of coins on the top. (Hammond 1963, 1957)

The allegation of fraudulent banking practices and insolvency of Illinois banks was that they had little capital reserves or liquidity to support their liabilities. Liquidity in particular requires a study of the portfolio of these banks. Many of the details stated in what follows rely upon Samad (1991) to which the interested reader is referred.

Review of the Literature

The study of the free banking period has been an important topic in the context of our present banking crisis and several studies have looked into our banking heritage. Hugh Rockoff’s pioneering thesis (Rockoff 1972) asserted that the provision of par value of the bonds, particularly when the par value of the bond was greater than the market value, caused wildcat banking in most states and thus failure of free banking. According to him, the states that witnessed wildcat banking were Illinois, Indiana and Wisconsin. Recent studies (Economopoulos 1990; Rolnick and Weber 1982, 1984; and Samad 1991) found no such evidence, for Illinois in particular.

Hilderlith and Rockoff (1973) made an important study on free bank's portfolio management which found the reserve position of the free banks in Boston, New York, and Philadelphia to be the product of careful decision making instead of erratic, gambling behavior. For example, risk-averting banks with a high capitalization rate were found to have held less reserve compared to banks with a low capitalization rate. Banks issuing more deposit liabilities as a source of bank funds were found to have held more reserves since they were exposed to higher risk than the banks which issued less deposit liabilities.

Rolnick and Weber's (1982, 1984, and 1986) studies provided new findings which challenged the traditional view of free banking failure. According to their studies, it was not Rockoff's par valuation provision but rather it was the sharp decline in the asset prices which were backing notes that caused the liability holders to panic and ultimately led to free bank failure. Their study covered four states, New York, Minnesota, Michigan, and Indiana. Illinois was not covered in their study nor did they study the portfolio management behavior of the free banks.

Economopoulos (1988, 1990) made two important studies. One study (1990) was related to Wisconsin Free Bank. In his study of Illinois Free Bank, Economopoulos (1988) examined two hypotheses: (a) the wildcat banking hypothesis and (b) the falling asset hypothesis. The main focus of his study was directed to testing the hypothesis that falling bond prices caused bank failures in Illinois, and he found some evidence in support of this. During the period of June 1860 and June 1861 he found many bank failures (89 banks failed) associated with the sharp decline—55 percent drop—of bond prices. “[T]hus the evidence indicates that the fall of bond prices was a major factor in the free bank failures of Illinois” (p. 263). He tested the wildcat banking hypothesis with the criteria of Weber and Rolnick and found that “traditional accounts of wildcat banking in Illinois are unsubstantiated” (p. 263). Economopoulos (1990) examined the portfolio management behavior of Wisconsin's free banks. He found that “solvent banks held more loans and specie, and issued more deposits and less banknotes than closed banks.” However the portfolio behavior of Illinois banks was not covered by either study.

Our study differs from the earlier studies in examining Illinois's wildcat banking through various indices to measure the bank's ability to pay short-term debt obligation and the bank's

way of utilizing debt (lender's money). These ratios, computed from the various aggregate balance sheets of the free banks, are used as the quantitative measure for testing the bank's fraudulent and liquidity behavior. Furthermore, we have made a comparative study with the portfolio balances of the Ohio free banking system. Ohio was one of the states which had a successful free banking system. It is, therefore, important to compare the portfolio behavior of the two banking systems to determine whether or to what extent Illinois experienced responsible banking practices.

Our study relies on different data sources. In addition to the State Auditor's reports, the most important source of Economopoulos's study, we have used various U.S. House Executive Documents and reports of the U.S. Controller of Currency. In fact, the latter two sources constitute the most important sources of our data base. Our results are similar to that of Economopoulos except in minor variation of number and the percentage, and thus serve to reinforce his conclusions on wildcat banking.

Therefore, the main focus of this paper is to examine the portfolio behavior of the free banks of Illinois in connection with the historical allegation.

Portfolio of the Illinois Free Banks and Evidences

Like any other business corporation, the Illinois bank of the antebellum period was a private firm that provided depository as well as note-issuing services. The banks were engaged in manufacturing and retailing "money" in their quest for profits subject to some constraints. While the banks held deposits from the individuals and businesses and issued notes to support their loans, banks were required to deposit an equivalent amount of the U.S. or state bonds to the State Auditor. These bonds were valued at the market price, *not* at their face value.

Their lending policy was also subject to some legal restrictions. The bank could not legally charge more than seven percent although the market interest rate was higher than seven percent (Cole 1919; Samad 1991). Here is a quote from Article 28 of the Illinois Law 1851:

That any such association or banker, in doing business under the provision of the Act, shall not be authorized to take or receive exceeding seven percent per annum as interest on real or personal security.

The bank's objective was to increase its shareholders' wealth. Profit maximization depends on successful portfolio management. Striking a balance between profitability and liquidity is the key concept for successful bank management. A prudent banker must choose assets, including liquid reserves, to ensure adequate protection from illiquidity while generating enough income to stay in business. This is not an easy task. Given a banker's preference for risk, portfolio choice becomes a function of the structure of asset and liability influencing risk and return. Although bankers' risk preference differed across banks, the bankers of Illinois free banks, in general, are assumed to have been risk averse instead of risk lovers.

Unlike today, there were no reserve requirements in the free banking era. As a result, a bank's own assessment of risk and return decided the portion of the total assets to be held in reserves. The banks of this period could issue deposit and note liability, apparently, without any restriction of reserve requirements. However, banks were required to deposit 100 percent U.S. or state bonds to the State Auditor as collateral to support their banknotes. Article 2 of the Free Banking Act of Illinois states:

whenever any person, or association of persons, formed for the purpose of banking under the provisions of this act shall lawfully transfer to and deposit with auditor any portion of public stocks issued, or to be issued, . . . such a person, or association of persons, shall be entitled to receive from the auditor *equal amount of such circulating notes.* (Illinois Law 1852)

The *Banker's Magazine* states:

when a general banking law was passed, which authorized any person or persons, on depositing with the Auditor of the State any of the stocks of the United States . . . *to receive an equal amount of banknotes, to be used for banking purposes.* (1854, p. 13)

Thus as long as a bank could provide U.S. or State bonds as collateral for the amount of note-liabilities, there was no restriction on a bank to issue notes. Illinois free banks were subject to much harder restrictions than the present commercial banks; like the demand deposits, the notes of the free bank issued were required to be redeemed in specie (gold and silver coin) on demand whenever they were presented to the counter of the bank. If a bank failed to redeem its notes on demand, the note-holder

had a legal right to sue and close the bank. Such a right of closing a bank lent an added emphasis for liquid reserves in the portfolio management of the free banking system.

The study of free banks' reserves, and their liquidity in particular, must be analyzed from the context of the economic environment in which the banks operated and the aggregate structure of their assets and liabilities.

Illinois Economic Conditions

An understanding of Illinois economic conditions in the 1850s—the structure of its population, agriculture, industry and transport—is essential for the study of banks' portfolio management, and lending behavior in particular. In the 1850s Illinois was making a rapid transition. It attracted a special breed of migrants during the second quarter of the nineteenth century. Whereas the earlier migrants of Illinois were, as Governor Ford described, “unambitious of wealth and great lover of ease,” the new migrants, “Yankees” in particular, who moved from the eastern provinces were great lovers of unending wealth and risk. They were talented workers, capitalists and above all entrepreneurs (Ford 1942; Pooley 1908). They were extremely eager for bank credit to build up their fortune.

During the 1830s and 1840s there had been tremendous development of river, canal, and railroad transport in Illinois. This transport development paved the way for the development of agriculture, industry, and commerce. The prairie land of Illinois was brought under large and commercial farming. Investments in agricultural implements and machinery demanded the release of bank credit to help sustain the growth of agriculture. The surplus produce of commercial farming had to find markets. People involved in marketing needed capital. Bank credit was the only source which could provide financial support to their needs.

Thus the need for capital, bank credit in particular, was fundamental to the farmers, manufacturers, and merchants. The state desperately needed bank credits to meet the economic needs of the people at the time. The market for bank credits already existed. Only local banks and their supply of bank credits were absent. “After the general crash in 1837, the state (Illinois) was without banking associations until 1851” (*Banker's Magazine*, July 1854, p. 13).

Some businesses from outside Illinois were taking advantage of this tight financial situation. For example, the Wisconsin Marine & Fire Insurance Company was successfully issuing notes

in Northern Illinois although it was declared illegal for Wisconsin Marine & Fire to issue notes. Although its notes were illegal, the people of Illinois who were in need of money were accepting them in large numbers. They paid no heed to the law.

The banks of Missouri and Ohio were also successfully banking in Illinois. Each year these banks earned substantial profits and transferred them from Illinois into their mother states. People in need of credit to meet their needs were accepting them without giving any consideration of the origin or legality of these banks.

Illinois free banks opened at a time when there was such a pressing need for their loans. In this condition of financial hunger, the supply of loans by way of banknote would not be turned down by individuals and would not be returned to the clearing-house for redemption (Selgin 1988, p. 67). The issue of banknotes to finance the credit needs of the individual and business acted as a medium of exchange side by side with specie.

Asset-Liability Structure of the Illinois Free Banks

The free banks of Illinois were required to report their assets and liabilities, the names of their stockholders and the bank's location to the Office of the State Auditor, and the auditor's office in turn published this information in its biannual reports. An aggregate balance sheet is randomly selected from the Bank Commissioner's Condition Report and presented in Table 1 to show the nature of assets and liability of the free banks.

Liabilities

It appears from the aggregate balance sheets (Samad 1991, various House Executive Documents) that the free banks of Illinois were engaged in issuing two major types of liabilities: banknote and deposits. The paid up capital of the banks was the major source of the bank funds except in three years, i.e., 1855, 1859, and 1860. The paid up capital was required to be invested in long-term U.S. and state bonds. By depositing these eligible assets, the banks obtained an equivalent amount of banknotes for circulation. These notes were exchanged for loans and discounts, for specie or additional investments in bonds, i.e., they acted as a medium of exchange. However, these notes constituted about 43 percent of the total liabilities. The year-by-year distribution of banknotes is presented in the Table 1 in Appendix 1.

The issue of notes had two special characteristics. Banks were exposed to high risk when the notes of the bank could not be

redeemed on demand. Failure to pay note-holders on demand would, in normal circumstances, result in the forfeiture of banking license. Such a harsh penalty provision provided bankers, prudent bankers in particular, an incentive to maintain adequate cash (specie) reserves or near cash reserve (which could be redeemed into specie with less time and cost) to meet the liquidity needs of the bankers.

The amount of banknotes in circulation was linked to the deposits of U.S. and states bonds. As such, the notes were safe and exchanged for loans and discounts or specie or additional investments in bonds. These notes acted as a medium of exchange just like specie.

As depository institutions, free banks were also issuing deposit liabilities. Deposits of the Illinois banks constituted a small percentage of the total liability because of the agrarian economic character of Illinois. Deposits were mostly savings. Deposit liability has advantages. One of the greatest benefits of issuing deposit liability is that the deposit holders could not invoke banking license on default. Thus, banks were not required to back these deposits with bond collateral. Secondly, interest payment on deposit was not a common practice during those days (Economopoulos 1990, p. 426).

In case of bank failure, note-holders' claim was applied first. This first lien provision and the legal rule that noteholders could revoke banking license if a banknote was not redeemed on demand provided an incentive for the bank depositors to put their money on saving deposit and monitor the financial position of the bank.

Miscellaneous liabilities constituted a very small percentage and therefore did not warrant discussion.

Assets

The assets of the Illinois free banks were made up of reserves and investments in varying combinations.

Reserves were the bank's most liquid assets and were mostly non-earning assets. Since reserve items are not profit making, banks usually keep them at a minimal level. As opposed to the division of reserves of the present time like legal reserve, required reserve, and excess reserves, etc., the free bank was not subject to any such legal reserve requirement. Banks decided the amount of minimum reserves as the amount which provided safety for the bank, on one hand, and maximum opportunity of earning on the other. The most liquid assets of the free banks

were, in descending order of liquidity, specie reserves, notes of other banks, bankers' deposits with other banks, additional investment in U.S. and states bonds, and real estates.

Specie in the bank's vault was the most liquid asset of the bank. It was the legal tender of the period. As the specie reserve was a non-earning asset, banks determined their amount based on past experience. The specie reserves of Illinois free banks against banknotes in circulation and deposits issued were, on average, 11 and 43 percent respectively.

In order to supplement specie reserves, the free banks of Illinois invested in the banknotes of the other banks. Notes of other banks in a bank's vault could be considered as near liquid. They could be cashed into specie as there existed markets for banknotes in New York, Philadelphia, and St. Louis. Investment in the notes of other banks minimized the need of excess specie reserves arising from note circulation. Investment in banknotes was, on average, seven and a half percent of the total circulation plus deposits. This procedure provided mutual insurance and reinforced credibility (Calomiris and Kahn 1990).

Illinois free banks were also found to have maintained investments with other banks. These investments were in the form of keeping deposits with other banks and buying the bills of other banks. Small bankers and county bankers traditionally maintained correspondent funds with the large and city banks so that whenever a small bank or a county bank was in need of extra specie, they could turn on them. The earnings from these investments were not the principal goal of the bank. It minimized specie needs and provided additional sources of reserves and thus ensured safety. Deposits with other banks and debt owing to banks other than loans and discount constituted about 30 percent of the total assets. Investment in others' banknotes plus bankers' deposit constituted about 43.8 percent of Illinois circulation and deposits.

The free banks of Illinois were found to have held other types of assets in their portfolios. In descending order of liquidity, the assets were compulsory investments in the U.S. and the states bonds and loans and discounts. The investments in loans and discounts and government securities were a bank's principal earning assets.

Investments in the long term government bonds constituted a significant portion of the total assets. It amounted to 100 percent of the note liabilities and about 48 percent of the total assets. This high percentage resulted from a number of reasons. *First*, the U.S. and States bonds were relatively the safest asset

to invest even though there were fluctuations of bond prices. *Second*, there was a ready market for bonds. Banks could easily sell their excess bonds without substantial loss in value. However, the securities deposited as a collateral to support banknotes were not allowed to be sold even when a dollar banknote was protested. *Third*, the rate of interest earned on bonds was relatively high and risk free. The default risk of government securities was almost nil.

The most unusual characteristic of bond investment was that although the free banks of Illinois were holding a large quantity of assets in their portfolio, *they were not permitted to use this asset/resource to meet any emergency of the bank*. They could be withdrawn and used only when the banks had met all obligations of the liability holders. This leads us into a central contention—free banks were virtually 100 percent reserve banks; their failure was a failure of regulation.

Loans and discounts were an important asset, accounting for about seven percent of the portfolio of the Illinois free banks. The banks could easily meet the liquidity needs by calling in some loans. Since they were short-term investments, the costs of converting them into specie were presumably not high. While they were assured of easily obtainable ready cash with less transaction cost, these investments were a good source of earnings when bills were bought and discounted.

Thus the Illinois free banks' supply of liquid reserves was defined as the combination of specie and demand claims on specie consisting of notes of the other banks, correspondent funds (which each bank held with other banks). In addition to these reserves, banks held a significant portion of their assets in loans, bills discounted, and other assets. Some of these assets were quite close to liquidity because any asset which could be converted into specie quickly and at a low cost could have potential ability to meet banks' liquidity. Although secondary markets for bank loans did not exist, they themselves were short-term ranging from three to twelve months (Economopoulos 1990). Moreover, these loans could be called in at short notice. Therefore, the free bankers could match the maturity of loan portfolio with the anticipated liquidity needs of the banks.

Comparison of Ratios and Evidence for Illinois

The study of the aggregate balance sheet revealed that the free banks of Illinois held large quantities of liquid, near liquid, and illiquid asset in their portfolios to support the total liabilities.

Table 1
A Summary of Comparative Figure of Ratios

Ratios	Illinois	Ohio
Capital Circulation + Deposit	.75	.48
Capital Circulation	.97	.69
Specie Circulation + Deposit	.097	.15
Others' Notes + Bankers' Deposit Circulation + Deposit	.34	.438
Capital Assets	.40	.28
Notes Circulated Deposits	.066	1.9

Whether these reserves were adequate for the honest portfolio management, given the structure of risk and return of their portfolios and Illinois economic condition, has been studied with reference to those of the Ohio banking system of the same period. The Ohio free banking system was chosen because the Ohio banking is unquestionably acclaimed as one of the most successful of the period. The various financial ratios like capital to deposit, capital to major liability, capital to asset, specie to notes, and specie to major liabilities provide deep insight into the operation of the banks portfolio management.

These ratios are used to measure the risk-return attitude of Illinois free bankers. The comparison of these variables between the Ohio and Illinois banking system would suggest whether or to what extent the behavior of the Illinois bankers was fraudulent. A summary of these ratios are presented in Table 1. The detailed year-by-year figures for these ratios would be available upon request.

Analysis and Interpretation of the Ratios

Capital Ratios

The total capital of a bank in itself has little meaning. However, capital can be compared more significantly to other items of

the balance sheet. Capital liability ratio is an important index of measuring a banker's preference for risk and return. The ratio of paid up capital to total liabilities provides a good measure of the proportion of bank funds not subject to sudden and unexpected withdrawal. Because of the stability of equity fund it was assumed that the higher the capitalization rate (i.e., the high capital-liability ratio) the lower the risk faced by a bank and vice-versa, *ceteris paribus*. The higher ratio of paid up capital to liabilities of a bank provides more protection to the liability holders than a bank with a small capital-liability ratio. Bank capital was intended to protect liability holders. Therefore, a risk-averse manager would hold smaller specie reserve if the capitalization rate was high. Hilderlith and Rockoff (1988) found the validity of this hypothesis for the banks of Boston, New York and Philadelphia.

The comparison of capital-liability ratio of the Illinois and Ohio banks presented in Table 1 shows that Illinois bankers maintained a significantly higher amount of paid up capital than that of Ohio bankers. The ratio of capital to major liabilities in Illinois and Ohio was, on average, .75 and .48 respectively. This high capitalization rate suggests that the bankers in Illinois were risk averse and their portfolio management was not fraudulent. Instead it was a product of careful decision making. They were not established to cheat depositors and noteholders.

Capital to banknote Ratio

Historically, regulatory authorities required banks to have total capital equal to about one-tenth of the total deposits to help protect the depositors. During the free-banking period, the main concern of the bankers was not the depositors but the noteholders. Because a bank's failure to redeem these notes on demand could close a bank. Therefore, the high ratio of capital to banknotes would reflect a bank's genuine concern for the noteholders. The comparative ratio of capital to circulation of Illinois and Ohio banks presented in Table 1 was .97 and .69 respectively. The higher ratio of capital to circulation would suggest that Illinois bankers did care for their liability holders and were risk averse. Capital represented those funds that a bank could lose without endangering its ability to pay its creditors. Compared to present time ten percent rule, Illinois bankers followed about 100 percent rule to protect their liability holders. So, the allegation of Illinois bankers' fraudulent behavior seemed unfair.

Specie Ratio

High ratio of specie to major liabilities, i.e., banknote plus deposits would indicate higher liquidity, therefore, lower liquidity risk exposure than a low ratio of specie to liabilities. The comparative specie ratios of the Illinois free banks and the Ohio free banks presented in Table 1 show that Illinois and Ohio bankers maintained specie reserves, on average, ten percent and fifteen percent respectively against their banknotes and deposits. Thus, it appears that the Ohio free banks' portfolios maintained five percent more vault cash/liquid assets than those of Illinois free banks. This would not mean that Illinois free banks were less efficient or fraudulent minded and that they would be exposed to high risk. It is consistent with the sound banking practice of the time. Deposits were a more volatile and riskier component of liabilities than notes (Myers 1931). Since Illinois banks issued more notes than deposits compared to those of Ohio banks it was consistent for Illinois bankers to hold a smaller proportion of specie reserves than that of Ohio banks. Hilderlith and Rockoff's study (1973) already supported the hypothesis that banks issuing more notes than deposits would hold less specie reserves in its vault.

The high capitalization rate and the high near liquid reserves of Illinois bankers would justify lower specie reserve in their portfolio. They held a significant portion of their assets locked in bond reserves. Moreover, Illinois free banks held a higher percentage of near liquid assets like notes of other banks plus bankers' deposits in their portfolio than that of Ohio free banks. The comparative ratio of notes from other banks and bankers' deposits to circulation plus deposits in Illinois and Ohio was 44 percent and 34 percent respectively. On the other hand, the five percent higher specie reserve by the Ohio free bankers would be expected because the Ohio banks held less capital against their circulation than the free banks of Illinois. Deposits of the Ohio banks as a source of bank funds were larger than notes in circulation. The deposits, as noted earlier, were more risky than notes.

Circulation Deposit Ration

A bank relying heavily on deposit as a source of bank's funds would be expected to bear greater risk than the bank relying on notes as a source of bank funds. Notes were issued to support loans and the loans had many stringent conditions attached to them. For example, borrowers receiving loans could not redeem

these notes within some specified time. Banks preferred to issue notes for the purchase of long distance goods and services (Golembe 1978). Banks discouraged loans by charging a high rate of interest if the note supporting loans were expected to return soon. In such loan markets, maintaining a high ratio of circulation to deposit would mean that the banker felt less at risk than the banker holding low circulation to deposit ration. Consequently, a bank with high ratio of circulation to deposit would hold less specie and more secondary reserves like notes of other banks, banker's deposits, etc. This policy would ensure safety as well as profitability. The comparative ratio of circulation to deposit of Illinois and Ohio banking system presented in Table 1 was .066 and 1.9 respectively. Since Illinois bankers were maintaining high ratio of circulation to deposit, it was expected that they would hold less specie and more secondary reserves which was reflected in the high ratio of other banks' notes and bankers' deposits to major liabilities (i.e., circulation and deposit). The ratio of secondary reserves in Illinois was .340 as opposed to .438 in Ohio shown in Table 1.

Structural Comparison Between the two Banking Systems

There have been some important structural differences between the Ohio and the Illinois free banking system.

In Illinois, there was only one type of bank, the free bank, in operation during the period under consideration. Consequently, they had no experienced model bank in the state to follow with respect to their lending and specie reserve policies. With the closing of the State Bank of Illinois in 1842, there was no other bank in the state than the Free Bank.

In Ohio, there had been three types of banks simultaneously in operation. They were: (1) branches of the Ohio State Bank, (2) Independent Bank and (3) Free Bank. Among these types of banks, the Ohio state banking system was experienced and constituted the largest share in the market in terms of bank number and bank capital in Ohio. The number of free banks and their total bank capital never exceeded one-third of the total Ohio State banks and their capital. Tables II and III substantiate our claim.

Since the Ohio State banking system had long been in operation prior to Ohio free banking, the Ohio Free Bank might have used them as their model in their banking practices. Consequently, the Ohio free banks as newcomer followed cautious banking practices with respect to specie reserve and lending policies while competing with the established banking firms.

Table 2
The Types of Banks and Their Total Number

Year	Independent Bank	Ohio State Bank	Free Bank
1853	3	44	13
1854	12	43	14
1856	12	42	13
1858	9	36	10
1858, Nov	9	36	11
1861	7	36	12
1862	7	36	12
1863	5	36	10

Numerous Free Bank Failures?

It was alleged that there were numerous free bank failures in Illinois. The view that there were numerous bank failures and that banks were in existence only for a short while during the free banking period were examined by Rolnick and Weber (1984) in four states: Indiana, New York, Michigan, and Minnesota, to test Rockoff's par valuation hypothesis (Rockoff 1974) and its impact on the bank's business life. This paper examines the bank failure rate and average bank life for Illinois free-banking period.

Bank failure and bank closing are not synonymous. When a bank paid all of its liabilities and then closed the business, it was closing of a bank at par. The closing of bank at par is not a failure of bank. It was a voluntary closing. A bank just like any other business might have closed banking firms voluntarily for different reasons. The closing of a bank at par involved no losses to the liability holders. So, the bank which closed at par could not be considered a failure.

A bank is said to have failed when it was closed below par. If the note-holders of a bank were not paid in full, then the closing of bank was considered a failure because the bank was ordered to be closed; it was an involuntary closing. Taking this criteria into consideration, Samad's study (1991) reveals that only two

Table 3
Capital and Asset Structure
of Free Banks vs. Ohio State Bank

Year	Ratio	Free Banks	Ohio State Bank
1853	Circulation/specie	2.38	4.36
	Liability/capital	4.50	4.00
1858	Circulation/specie	3.35	3.72
	Liability/capital	4.12	3.14
1861	Circulation/specie	3.38	3.71
	Liability/capital	3.29	5.09
1863	Circulation/specie	1.47	3.44
	Liability/capital	5.88	4.45

banks failed before 1861 although a total of 25 banks were closed. The failure rate is only 1.5 percent in nine years. This 1.5 percent failure rate (or two failed banks) during 1852–1860 speaks volumes for it does not mean there were numerous bank failures nor does it mean an unstable banking system. Free-bank failures however jumped up in 1861 when 94 banks closed. Of these, 89 banks were found to have failed. The failure rate in a single year, the first year of the Civil War, was abnormally high in Illinois. The failure rate was almost 95 percent in a single year. The large bank failure in a single abnormal year could not decide whether the free banking system of Illinois was a failure or success. If the failure rate of 1861 is ignored, the free-banking system is a relatively stable system. The total number of banks failing during the period 1852–1861 and 1862–1864 was only three although a total of 43 banks closed during the same period. After the initial shock of the Civil War, bank failures almost ceased to exist in Illinois. One bank was found to have failed between 1862 and 1863.

The issue of the free bank's numerous failures and short existence of operation explained by Rockoff is carefully examined in Illinois. According to Rockoff, it was profitable for a bank to close its business permanently when market value of the bonds backing the notes as collateral was less than bond's par value (face value) because notes were issued to the amount of the face value of the bonds. Thus, as soon as the circulation of notes was

complete, banks found it profitable to close the business. The life span of the banks in such a situation was found to have been very short—"maybe a month or two" (Rolnick and Weber 1982).

Rolnick and Weber's method is used for calculating a bank's average life. A bank's business life is counted as the difference between the date a bank was closed and the date the bank was opened for business. Using this technique, the estimated average life span of Illinois free banks does not substantiate the conventional view that Illinois free banks opened their business just for a short while to circulate their notes. This study finds that the average life span of an Illinois free bank was around three years and eight months. Only 20 or 14 percent of the banks were in operation for a period of less than one year; 29 percent operated between one and two years; 13 percent operated between two to three years; and the remaining 57 percent of the banks were in business for a period of more than three years. This is somewhat better than that noted by Economopoulos (1988).

This estimate of the average bank's life provided in this study is conservative. Each bank's opening date is counted from the date the bank's name was found to have first appeared in the condition report of the Bank Commissioner. When a bank's name was not found in the current condition report, then the date of the previous condition report showing the bank's name was considered as the closing date for that bank.

In such an estimate there is downward bias in the bank's average life span. Many banks could have been in operation for one to six months before they were found to have appeared in the bank's condition report. Similarly, many banks could have been in operation for a few more months before they had been reported in the next condition report. As such, my estimate of Illinois free banks' average years of operation is downward biased in both respects.

So, the allegation of instability of the Illinois free banks characterized by numerous bank failures and the short existence of banks' business operation is an exaggeration.

The allegation of banks' unfair practices by locating their offices in "an unknown and inaccessible place" was also studied. Only four banks out of 141 were found to have no exact place of location. These banks were: the Bank of Identity, the Bank of Kewanee, the Eagle Bank of Illinois, and Canal Bank. Their location could not be found in the 7th and 8th U.S. census. This meant that these banks were either fictitious or they were in the remote countryside where population figures were too low to

warrant census attention. The rest of the banks of Illinois (98 percent) were found to have been established in the county seats and in important townships. Judged against the location of population, our study shows a large majority of the banks were in communities of more than 1,000 people (Quinn and Samad 1991).

Thus, a generalized allegation of free banks' involvement in fraudulent practices is unjust and unfair.¹

Conclusions

According to aggregate balance sheets, the free banks of Illinois held wide varieties of assets in their portfolios and the value of these assets were larger than those of their liabilities. Therefore, the banks were never insolvent. The banks might have become temporarily illiquid but they were found to have held huge amounts of other resources in their portfolios. Whenever the value of their liabilities were exceeding the value of their assets, banks were issued margin calls or required to retire some of their notes from circulation. Their portfolio management was, thus, not intended to cheat the public. The aggregate balance sheet could not provide any conclusive evidence in support of the conventional view that the portfolio management of the free-banking system was fraudulent. It does not, however, rule out the possibility of irresponsible banking. There might have been some banks that were established with the fraudulent intention of circulating notes but not redeeming them.

The literature on free banking has laid inadequate emphasis on the financial hunger that characterized many states and which forms an appropriate background against which the responsibility of banking must be judged. The Bank of the State of Missouri wished to set an example of financial integrity in 1838 by refusing to issue its notes unless it covered them securely. Such was the demand for finance that a committee of 11 prominent citizens offered to endorse paper from the banks in Illinois and thus guarantee the Missouri Bank from loss. The Missouri Bank refused. In disgust, the merchants and traders of St. Louis moved increasingly to private banks and insurance companies (Hubbard and Davids, p. 60).

¹The approaching Civil War caused panicky noteholders to demand redemption and cause the price of Southern bonds to decline by 55 percent between 1860 and 1861. This contrasts with drops of 13 percent during the panic of 1854 and 24 percent during the panic of 1857. As Southern bonds were a major source of collateral for Illinois free banks, the subsequent failures became more understandable. A detailed study of failure due to the Civil War is underway and will be reported subsequently.

The center of economic activity moved from St. Louis to Chicago in the period 1840–1860 (see the maps in Cronon, *Nature's Metropolis*, chap. 6), so we can presume the pressure on Illinois banks to be just as great. That the Illinois banks were not principally, or even largely, wildcat banks was adequately shown by Economopoulos earlier by looking at location and at bond prices. (The only possible weakness being the focus on Tennessee bond prices, when Missouri bond prices would have been more relevant.) The Illinois banks had to issue notes at the market price of their bonds—thus providing for 100 percent reserve. That regulation was at least somewhat responsible for the failures of these banks, and was evidenced by the bank commissioners, who protested the notes not the noteholders and from a defective scheme whereby Illinois bonds, which everyone would be most cognizant of, were held at only 80 percent of their value while the bonds of other states could be held at full value. A close look at the balance sheets shows most banks did hold full reserves and therefore can be said to have been free and responsible banks.²

²An anonymous referee pointed out that this outcome is consistent with the prediction of Ludwig von Mises, *Human Action* 3rd ed. (Chicago: Henry Regnery, 1966), pp. 444–48.

Appendix
Statement of the Condition of the Banks
of the State of Illinois, on Monday, April 3, 1854

Resources

Total public stocks at which they were received by the auditor	\$2,475,741.62
Amount paid over the value at which they were received by the auditor	\$196,162.13
Real estate	\$ 3,158.22
Notes of the other bank on hand	\$385,339.45
Amount of debts owing to the associations other than loans and discounts	\$1,368,203.68
Suspended debt	
Specie on hand	\$565,152.04
Loans and discounts	\$316,641.76
Deposited with other banks	\$878,612.58
Expense accounts	\$4,874.97
Checks, drafts, and other cash items	\$63,892.41
Total resources	<u>\$6,257,778.86</u>

Liabilities

Capital stock paid in and invested according to law	\$2,513,790.17
Amount of debts owing by the association other than depositors	\$294,034.50
Amount due to the depositors	\$1,286,102.25
Notes or bills in circulation	\$2,283,526.00
Profit and loss account	\$71,787.00
Total liabilities	<u>\$6,449,239.92</u>

Source: U. S. House of Representatives, Executive Documents 33-2.

References

- Calomiris, C. and C. Kahn. 1990. "The Efficiency of Cooperative Interbank Relations: The Suffolk System." Working Paper: University of Illinois.
- Cole, Arthur C. 1919. *Constitutional Debates of 1847*. Springfield: Illinois State Historical Library.
- Cronon, William. 1991. *Nature's Metropolis*. New York: Norton.
- Gorton, Gary. 1985. "Banking Theory and Free Banking History. A Review Essay." *Journal of Monetary Economics* 16.
- Green, George D. 1979. "Louisiana 1804–1861." In *Banking and Economic Development*. Rondo Cameron, ed. New York: Oxford University Press.
- Economopoulos, Andrew J. 1988. "Illinois Free Banking." *Journal of Money, Credit, and Banking* 20: 249–64.
- . 1990. "Free Bank Failures in New York and Wisconsin: A Portfolio Analysis." *Explorations in Economic History* 27: 421–41.
- Ford, Thomas. 1942. *The History of Illinois as a State in 1818 to 1847*. Chicago: Lakeside Press.
- Golembe, H. Carter. 1978: *State Banks and the Economic Development of the West 1830–44*. New York: Arno Press.
- Hammond, Bray. 1957. *Banks and Politics in America from the Revolution to the Civil War*. Princeton: Princeton University Press.
- Hilderlith, H. Roger and Hugh Rockoff. 1973. "The Management of Reserves by Antebellum Banks in Eastern Financial Centers." *Explorations in Economic History* 2, No. 1.
- Hubbard, T. W. and L. E. Davids. 1969. *Banking in Mid-America*. Washington, D.C.: Public Affairs Press.
- Hunt's Merchant Magazine*, several issues 1852–1862.
- Illinois State Auditor Biannual Report*, Springfield, Ill.: 1852.
- Luckett, Dudley. 1980. *Money and Banking*. New York: McGraw-Hill.
- Myers, Margaret. 1931. *Origin and Development*, Vol. 1. New York: Columbia University Press.
- Poole, William V. 1908. *The Settlement of Illinois from 1830 to 1850*. Unpb. Thesis. Madison: University of Wisconsin.
- Rockoff, Hugh. 1974. "The Free Banking Era: A Re-Examination." *Journal of Money, Credit and Banking* 6: 141–67.
- . 1972. *The Free Banking Era: A Re-Examination*. New York: Arno Press.
- Rolnick, A. and W. E. Weber. 1984. "The Causes of Free Bank Failure: A Detailed Examination." *Journal of Monetary Economics* 14: 1–24.
- . 1982. "A New Explanation for Free Bank Failures." Minneapolis, Minn.: Federal Reserve Report 79.
- Samad, Abdus. 1990. *The Free Banking Experience in Illinois*. Unpb. Ph.D. Dissertation: University of Illinois.
- , with Stephen Quinn. 1991. "Illinois Free Banks." *Illinois Business Review* 48 1: 3–9.

Selgin, George A. 1988. *The Theory of Free Banking*. Totowa, N.J.: Rowman and Littlefield.

U.S. House of Representatives, 33rd Cong., 1st Sess., Ex. Doc. 102.

U.S. House of Representatives, 33rd Cong., 2nd Sess., Ex. Doc. 82.

U.S. House of Representatives, 34th Cong., 1st Sess., Ex. Doc. 101.

U.S. House of Representatives, 34th Cong., 3rd Sess., Ex. Doc. 87.

U.S. House of Representatives, 35th Cong., 1st Sess., Ex. Doc. 105.

U.S. House of Representatives, 35th Cong., 2nd Sess., Ex. Doc. 112.

U.S. House of Representatives, 36th Cong., 1st Sess., Ex. Doc. 49.

U.S. House of Representatives, 36th Cong., 2nd Sess., Ex. Doc. 77.

U.S. House of Representatives, 37th Cong., 3rd Sess., Ex. Doc. 24.

U.S. House of Representatives, 38th Cong., 1st Sess., Ex. Docs. 12–26.

Hayek, Business Cycles and Fractional Reserve Banking: Continuing the De-Homogenization Process

Walter Block and Kenneth M. Garschina

Science sinks or swims based on the quality of the distinctions it makes, and social science is no exception to this general rule. It is as important to make accurate differentiations in the history of economic thought as it is in any other branch of this discipline.

In this regard, the accomplishments, writings, and analytic apparatus of Ludwig von Mises and his pupil and friend, F. A. von Hayek, have been widely viewed as all but indistinguishable. And this holds true not only within the profession as a whole, but also among economists associated with the Austrian or praxeological school.

There is good reason, at least at first glance, for such a conflation. Both economists shared, or at least appear to share, a philosophical outlook, and a methodology; their views on socialism, government regulation of the economy, the free society, and the causes of the business cycle, were in many ways similar. But there were also some sharp and important differences between them, which are rather technical. Perhaps this is one reason why they have been little appreciated. But these divergences are basic, with implications for the entire corpus of Austrian economics, and, indeed, economics in general. It is therefore all the more important to distinguish between the views of these two scholars.

Walter Block is associate professor of economics at the College of the Holy Cross. Kenneth M. Garschina graduated from the College of the Holy Cross in 1993. We would like to thank Murray N. Rothbard and 3 anonymous referees for helpful comments. The usual caveat applies.

The Review of Austrian Economics Vol. 9, No. 1 (1996): 77–94
ISSN 0889-3047

Salerno forcefully makes the point that the unrecognized incompatibility between Mises and Hayek is of far more than mere antiquarian interest:

Unfortunately, the majority of those who currently regard themselves as “Austrian economists” have failed to recognize the considerable differences between these two paradigms. And because Mises was the main influence on Hayek’s early writings on business cycle theory and on socialist calculation, the most important manifestation of this failure is the tendency to attribute to Mises positions originated by Hayek or independently developed by those working within the Hayekian paradigm. This tendency is reinforced by what may be called the “Whig presumption,” still inexplicably prevailing among many Austrians despite the publication of Thomas Kuhn’s book three decades ago, that since Hayek “came after” Mises he must have incorporated in his own work all that was worthwhile in his predecessor’s. The result is that attention has been deflected from the Misesian paradigm, and those seeking to deepen and extend it have found it increasingly difficult to gain recognition for their own efforts or to channel the interests and efforts of younger Austrian scholars into the same endeavor. There thus currently exists a pressing need, especially for Misesians, to undertake the task of a courageous and thoroughgoing doctrinal dehomogenization of Hayek and Mises. (Salerno 1993, pp. 115, 116)

It is not sufficient to show only that the perspectives of Mises and Hayek are not fully reconcilable; and that this fact is not widely appreciated. Once this is conceded, the question naturally arises, Which is correct and which is not? Therefore, it is important to follow Salerno’s lead even further, and take a stand on that issue as well.

There is a small but ever growing literature which might be called “Hayek revisionism.” It takes the view that the analysis of the teacher is very distinct from that of the student, and vastly preferable. Hayek, a 1974 Nobel Prize winner in economics, is widely known as a radical advocate of the Austrian or free enterprise philosophy. And to a certain extent this reputation is well deserved. After all, Hayek (1944, 1989) are classic critiques of socialism and central planning, Hayek (1960, 1973) defend the rule of law, Hayek (1978) shows the flaws in “indicative” or “market” planning, and many of his other books and articles demonstrate the

beneficial workings of the market (1948, 1954, 1967, 1981). Of late, however, scholars have shown that some of his most basic writings cannot be reconciled with a thorough going adherence to praxeological analysis (Salerno 1993) and economic freedom (Rothbard 1982).

Even within the corpus of Hayek's own work a distinction may be made. A scholar who distinguishes two different strains of thinking within Hayek's own writing was Hutchison (1981). He labels the early publications as Hayek I (before 1936) and the later ones as Hayek II (1937 and thereafter). Of the earlier period Hutchison (1981, p. 211) states: "Affinities with the ideas of Austrian predecessors, notably with those of his 'mentor' Mises, are apparent." In contrast, the first publication of the latter period (Hayek 1937),¹ Hutchison comments:

It certainly marks a vital turning point, or even U-turn, in Hayek's methodological ideas, *and ought to be, but has not been recognized as marking a fundamental shift* . . . The main insights of this article are quite incompatible . . . with the methodological ideas in his previous writings. (1981, p. 215; emphasis in the original).

The new dispensation in Hayek had mainly to do with a shift from praxeological (e.g., Misesian) methodology to that based on logical positivism (e.g., Popper), and from an emphasis on appraisalment to one of lack of full information regarding questions of central planning and socialism (Salerno 1993). This is not to say that in the earlier period Hayek was indistinguishable from Mises, nor that the latter period constituted a total break. There were differences before, and similarities afterward. But it is our contention that even though Hayek I was preferable to Hayek II, the errors in the former are still well worth exploring.

Following in Hutchison's footsteps on this research is Salerno (1993). Salerno has shown that as the years went by, and Hayek moved from his Hayek I position to his Hayek II views, he pulled further and further away from the uncompromising praxeological and free market analysis of his mentor Ludwig von Mises (1963); that whereas Hayek I was reasonably close to Mises in many ways, Hayek II began resembling him in philosophical outlook less and less.

¹For a good critique of this paper, see Selgin (1988, pp. 28, 29).

In the view of Salerno (1993),² there is not one Austrian strand emanating from Menger (1950), the founder of this School, but rather two. The first is transmitted to us by Böhm-Bawerk (1959) and Mises (1912, 1957, 1966, 1981); the second comes to us courtesy of Wieser (1967) and his follower Hayek. Salerno's contention, and our own, is that the first strand is preferable to the second (1993, especially footnotes 3 and 4). As well, and perhaps of even greater importance, Salerno shows that even the relatively preferable version, Hayek I, is not without its flaws. We shall try to show several of them: business cycles, fractional reserve banking, governmental growth enhancement, and 100 percent money.

Business Cycles

The majority of contemporary viewpoints within the economics profession favor a strong role for the state as necessary to combat the business cycle.³ With regard to the problem of booms and busts in particular, it is the consensus among economists (Frey et. al., 1984; Block and Walker 1988) that the market, uncontrolled by central authority, will continually veer into either unemployment or inflation.

In contrast, it is the Austrian contention that these problems are not "natural" results of the market system; on the contrary, they are in large part created by interventionistic acts on the part of the government in the first place. The public sector, in this view, is the problem, not the solution.

Hayek (1931) is clearly part of Hayek I. And not only that: it is also part of the Hayek I contribution which is not at all problematic. In it, he makes the point that our inability to tame market instability is not due to deficient economic acumen on the part of members of the private sector. Rather, it comes about because of the interference and regulation of credit markets by the state. Specifically, this follows from credit expansion, which drives interest rates down below the levels which would otherwise result. This, in turn, leads entrepreneurs to mistakenly

²Friedman (1991) made much the same point about the publications in the two epochs of this scholar's life, only in support of Hutchison, and in sharp contrast to Salerno, he praised Hayek II while denigrating Hayek I. Friedman's distinction, although based on methodological differences, mainly revolved around the psychological issue of "intolerance."

³For this view, which encompasses virtually all of the mainstream perspectives, see Friedman and Schwartz (1963), and Keynes (1936). For a reply to the latter, see Hoppe (1992).

invest in the higher orders of production. But Hayek is careful to point out that the error is only from the long term point of view: in the immediate run, placing money in heavy industry is fully justified by the now (artificially) lower rates of interest.

Hayek (1931) leans heavily on the work of Mises (1912, 1966); his, like his mentor's, is a malinvestment theory of depressions: these cycles come about not because of too much⁴ investment, nor yet because of too little. For all that can be known, exactly the "right amount" of investment may be undertaken. But because it enters too high in the structure of production, compared with where it would have gone had businessmen not been subsidized by low interest rates, the seeds of future economic destruction are sown. Moreover, in the Misesian tradition, Hayek (1931) makes important contributions of his own. For one thing, the now famous⁵ "Hayekian triangles" owe their appearance to this work.

If Hayek (1931) was a part of the Misesian Hayek I, then Hayek (1933) would have to be counted as an aspect of the non- or anti-Misesian⁶ Hayek I. In this discussion on cyclical fluctuations, he denies that banks are wholly or even partially responsible for the nature of the recurring trade cycle. He contends that these financial institutions have never been prohibited from holding fractional reserves and therefore should not be held responsible for any of the repercussions. We flow, seemingly endlessly, from periods of prosperity to periods of struggle and recuperation, but Hayek labels it "nonsensical" to blame banks or to hold any other party "guilty" for the continuous boom-and-bust nature of our economic cycle. Hayek states:

we can also see how nonsensical it is to formulate the question of the causation of cyclical fluctuations in terms of "guilt," and to

⁴This is not an "over-investment" theory, as asserted by Hutchison (1981, p. 224, n. 1); rather, strictly speaking, it is a misallocation of investment theory. This is a fine point of distinction, as for the Austrians there is typically excessive investment in the boom phase of the cycle. But this is not logically necessary. The key point is that these funds go to orders of production which are too high (e.g., far removed in time from consumption) for sustainability, given the public's time preference rates. Although unusual, it is compatible with the praxeological theory of the cycle to contemplate *under* investment as a causal agent. As long as these funds are placed in improperly high orders of production, there can still be projects begun for which financing necessary for completion will not be forthcoming.

⁵If only within Austrian circles.

⁶A way of characterizing this, alternative to Hutchison's (1981), is that we date the onset of Hayek II earlier than Hayek (1937). For us, it occurred at least as early as Hayek (1933).

single out, e.g., the banks as those “guilty” of causing fluctuations in economic development. Nobody has ever asked them to pursue a policy other than that which, as we have seen, gives rise to cyclical fluctuation, seeing that the latter originate not from their policy but from the very nature of the modern organization of credit. (Hayek 1933, p. 189)

Now this is more than just passing curiosity. If true, it would be, perhaps, the first case on record in all of recorded economic history, where an industry took no interest whatsoever in the regulations pertaining to it, nor in proscribing competition against extant members.

On the face of it, it would be as if the taxi industry were completely unconcerned with legislation that limited the participation of gypsy cabs (Williams 1982, chap. 6), or as if the American Medical Association were totally uninvolved in precluding the entry of new doctors into that profession (Friedman 1962, chap. 9; Hamowy 1984). In perhaps the most famous statement in all of economics, Adam Smith warned that:

People of the same trade seldom meet together, even for merri-
ment and diversion, but the conversation ends in a conspiracy
against the public, or in some contrivance to raise prices. (1776,
vol. 1, bk. 1, chap. 2)

According to the view of Hayek we are now considering, bankers, of all people, would appear to be an exception to that general rule.

Fortunately, we need not rely on theoretical public choice (Buchanan and Tullock 1971; Buchanan, Tollison, and Tullock 1980) and realistic historical investigation (Kolko 1963) to show that Hayek’s belief is without merit. There is also a plethora of empirical examples which can serve as a refutation of the banker-as-innocent hypothesis.

For example, Paul and Lehrman note that

America’s bankers had long chafed to cartelize the banking system still further. . . . The growing consensus [in the nineteenth century], then, was to redirect the banking system by establishing, at long last, a central bank. The central bank would have an absolute monopoly of the note issue, and reserve requirements would then ensure a multilayered pyramiding on top of these central bank notes, which could bail out banks in trouble, and, moreover, could inflate the currency in a smooth, controlled, and

uniform manner throughout the nation.⁷ (Paul and Lehrman 1982, pp. 119–20)

There is another grave problem with Hayek's 1933 analysis. He believes that the banks are not "guilty" of causing business cycles *also* because he thinks that in the early stages the "natural rate of interest" or profit on the market increases, and that the banks are not astute enough to realize it, so that they only pull the loan rate of interest below the natural rate, that is, by not raising their loan rates fast enough to match changes in the natural rate. The difficulty with this is that it misconceives the Misesian (1912) insight. The problem is *not* one of omission, rather it is one of commission; it is not that the banks are too passive and ignorant about finding the right loan rate to match the natural rate. Instead, it is that they actively *expand credit* beyond the cash in their vaults, thereby *pushing* the loan rate below the natural rate. In short, the Misesian view is that the banks don't have to search for the natural rate in order to avoid generating the business cycle; all they have to do is not expand credit beyond their cash holdings. This is surely a much easier task. The banks' insistence on expanding credit generates the business cycle, and makes them responsible and thus "guilty" as charged.⁸

Fractional Reserve Banking

But Hayek is not content to exonerate bankers as embodiments of free enterprise virtue. He goes on to offer a defense for their anti-market activities (the harm of which he has just finished denying). Hayek maintains:

So long as we make use of bank credit as a means of furthering economic development we shall have to put up with the resulting trade cycles. They are, in a sense, the price we pay for a speed of development exceeding that which people would voluntarily make possible through their savings, and which therefore has to be extorted from them. (1933, pp. 189–90)

During the course of his discussion, Hayek focuses on the structure of our current monetary organization of credit and upon the inherent flaws in this structure that create the cycle. He is

⁷For a critique of this initiative, see Hoppe (1993) and Rothbard (1975, 1994).

⁸The authors of this paper wish to thank Murray N. Rothbard for this insight.

correct in identifying fractional reserve banking in particular as a major source of disruption to economic welfare, but then fails to label the state's utilization of this system as detrimental to overall growth.

The most glaring manner in which government has impeded the natural workings of the free market lies with its ability to control the volume of money. Hayek made this the basis of his Misesian-based (1931) theory of cyclical fluctuations. With the introduction of money to a society, control over the economy can be shifted from the individual's natural tendency to produce and trade to one where government disturbs and hampers the production process through its manipulation of the money supply.

Changes in the volume of money by the government can be effected in two ways: alteration of note circulation by central banks and by "creation" of deposits in other banks. Hayek correctly argues that it is the ability of independent banks to "create money" that is harmful to the economy. But these financial institutions are able to increase the money supply due to the system of fractional reserve banking. For example, if a deposit is made of one hundred dollars, the bank is only required to hold "in reserve" or "on hand" a small fraction of this amount. The rest can be granted as credit to customers who will inevitably follow the same deposit process with their newly acquired funds. In this way, in a decentralized system, money travels from bank to bank, multiplying each time it is lent out. And the original depositor, of course, is still able to draw on the funds entrusted to the bank on demand. As the process continues, the volume of money increases, lowering the money rate of interest below the natural rate, which Hayek (1933, p. 147) defines as the rate "at which the demand for and the supply of savings are equal."

Rothbard (1975, p. 19) agrees with Hayek on his thesis with regard to the causes of cyclical fluctuations, and refers to a "boom" period as one of misinvestment created by the government sanctioned credit system through its control of the money rate of interest. With banks offering credit at an artificially low level of interest, capitalists invest in production processes that must inevitably be abandoned when the banks eventually curb the amount of credit being offered because of increasing cash requirements or a rise in the discount rate (Hayek 1933, p. 175). As soon as the banks cease to increase the volume of money in circulation, the interest rate at which credit is offered will rise to the natural level and leave unfinished the investments previously made possible by increased levels of credit. The freedom given to the banks

by the state to control the volume of money and interest rates initiates production that cannot possibly be completed. The periods that we know as “crisis” or “depression,” or, in the most recent euphemism, “recession,” are in fact the time needed for the process of abandoning or reallocating the investment mistakes of the boom period.

Despite his accuracy in identifying the source of fluctuations, Hayek suggests that we must continue to use fractional reserve banking in order to spread the development of technical and commercial knowledge. This, despite the price paid in economic disruption during every bust period. He states:

And even if it is a mistake—as the recurrence of the crises would demonstrate—to suppose that we can, in this way, overcome all obstacles standing in the way of progress, it is at least conceivable that the non-economic factors of progress, such as technical and commercial knowledge, are thereby benefited in a way which we should be reluctant to forego.⁹ (Hayek 1933, p. 190)

He contends that extension of credit, even though it results in a recurring crisis, is necessary in order to enhance man’s ability to discover and produce things not possible from his own personal savings. Hayek views the “benefit” derived from providing credit to those not in effect credit worthy as outweighing the consequences of decimating the entire economy every few years. But this is mistaken. It is in fact an undermining of Hayek’s own work and defeats the logic of his entire business cycle discussion.

Contrary to Hayek,¹⁰ in order to enhance economic welfare, any prospective technological or commercial advancement should be funded based upon its own merits, and not depend upon an artificially low money rate of interest. An economy void of fractional reserve banking would be less able to overextend itself through excess credit and more likely to produce an optimal amount of technical and commercial services. These businesses may not come to fruition as quickly and powerfully as they would were they backed by artificially extended credit, but the economic foundation predicated upon voluntary choice will be stronger. The

⁹Hayek’s reference is presumed by the present authors to be to “existing monetary organization” (1993, p. 187) of which fractional reserve banking is an integral part.

¹⁰That is, to the flawed part of Hayek I whom we have been citing.

percentage of failures, e.g., wasted resources, will be therefore reduced. Moreover, an economy that sustains constant growth will outproduce one which sacrifices an undetermined number of years to crisis in order to artificially encourage growth.

Nor is this a matter of mere cost-benefit analysis. The point here is not that of the two values, economic stability and technical progress, we hold that the former necessarily outweighs the latter. On the contrary, given the impermissibility of interpersonal comparisons of utility (Rothbard 1977), our view is that it is impossible, *a priori*, to determine which one is more important. Why, then, our opposition to Hayek's preference for technical progress *vis-à-vis* stability?¹¹ It is because the burden of proof is on him who would upset the natural order of the laissez-faire economic system, and Hayek has not even seen this as a challenge, let alone attempted to overcome it.

In order to see this point more clearly, suppose that someone, call him Mr. H, had contended that war enhances scientific innovation (radar, better planes, rockets, improved medical techniques learned on the battlefield). And that, further, the value of these improvements was greater than the loss due to people being killed in war. One possible response would be based on a cost benefit analysis. Here, we might make the contrary claim that no deaths due to battle impose more of a loss on humanity than the inventions thereby conferred gain for us. But interpersonal comparison of utility considerations render such a tack invalid. Instead, we would say that the natural order of society is peace, and that the intellectual burden of proof rests on those such as Mr. H who claim, somewhat paradoxically, that the human condition can be improved by fomenting armed hostilities. It is clear that this burden has not been upheld, indeed, nor can it be.

Governmental Growth Enhancement

Hayek (1933, p. 191) also speaks about the "utilization of new inventions and the realization of new combinations." He claims that they would be made more difficult in the absence of cyclical fluctuations, and that the psychological incentive towards progress would

¹¹We assume, if only for the sake of argument, that Hayek is correct in his contention that there is a tradeoff. That is, that one can indeed gain more by destabilizing the economy in the form of new and better inventions than by allowing it to proceed freely and (relatively) steadily. Unfortunately, Hayek gives us no good reasons to suppose that this form of government intervention will promote innovativeness.

be retarded.¹² But the very opposite is true. Namely, each year many businesses are not launched simply because of fear of crisis. Capitalists would be more inclined to utilize venture funds if relatively constant growth became an expected reality, for potential investors would not have to continually fear a business-crippling recession.¹³

More radically, Hayek's conception of an increased technological or commercial rate of progress is flawed in and of itself. By offering credit to those not deemed worthy of it by the market (Hazlitt 1946, pp. 30–40), we push ourselves beyond the scale of development for which the economy is ready. There is an optimal amount of forward movement that any economy can accommodate. To overshoot that appropriate level is to attempt to advance to a degree unmanageable by society and ultimately by the individual. There exists a natural order for the structure of production, whether in the realm of physical output or of scientific and technical ideas. If so, any compulsory attempt to exceed it is logically doomed to failure. At present, lending institutions are permitted to alter the path of growth through extension of credit. This not only gives impetus to the business cycle, it also cannot succeed in its self-avowed goal of increasing the rate of technical progress.

Kirzner speaks of this phenomenon in terms of:

an intertemporal equilibrium. Plans made today must fit not only with plans made by others today [intra-temporal equilibrium], but also with plans made in the past and other plans to be made in the future. A state of equilibrium will not exist wherever any plan being made at any date fails to dovetail with other relevant plans of whatever date in the entire system being considered. A man who erects a shoe factory and who discovers in later periods that shoe leather is unobtainable, or that consumers no longer wish to buy shoes, made his decision in ignorance of the plans of others on which his own depended. A man who educates himself in a profession for which later demand is lacking has made a plan based upon incorrectly anticipated plans of others. (Kirzner 1979, p. 112)

¹²Even if true, this does not satisfy the burden of proof incumbent on Hayek. For that to be achieved he would have to assert not only that booms and busts enhance economic innovativeness, but that it does so in a manner that *more than offsets* the attendant economic losses.

¹³We are of course in effect making the usual assumption about risk-avoiding preferences.

As Kirzner points out, it is indeed possible for entrepreneurs to act incompatibly with intertemporal equilibrium. When they do so in a market context, of course, they suffer the consequences, and, as a result, this sort of misallocation tends to be minimized.

However, as Hayek does not seem to appreciate, governments, too, can engage in intertemporal misallocation, and a paradigm case in point is an attempt on the part of the state to promote overoptimal economic and scientific development. At the outset, this sounds like a contradiction in terms. How, after all, is it possible to have *too much* economic growth? One possibility, furnished by Hayek himself, is governmental monetary policy which results in a below market rate of interest, which leads to basic investments which cannot be completed, e.g., the classical Austrian business cycle time misallocation of the structure of production (Rothbard 1975; Mises 1963; Hayek 1931).

Another example might well be President Nixon's "moon shot" of several decades ago. This was a "success" in that several taxpayers were indeed launched up to this celestial body, and made it back home all in one piece. But it is unlikely that this was an impetus to the overall goal of space exploration; it is more probable that it came too soon, before the complementary factors of production were in place. The point is, had the billions of dollars spent been used instead for research and development in fuels, rocketry, life support systems, human (scientific) capital, etc., it is entirely possible that the human race would have been, by now, far ahead of where it actually is in this regard.

It is thus not a matter of weighing additional economic growth against the ravages of the business cycle. The latter is, of course, deleterious—not only to "children and other living things"—but to the entire economy. The former, however, is *also* a denigration of economic welfare, and cannot, therefore, be considered as a positive offset to the admittedly harmful boom and bust cycle.

There is another way to make this point. The Hayek I who supports fractional reserve banking and government interference with the market in order to spur "growth" is an economist who is in effect calling upon the central banking system to determine the evenly rotating economy's interest rate. That this cannot be done is not due merely to a lack of knowledge, a continual refrain in the Hayekian *oeuvre* (Salerno 1993). The problem is, fractional reserve banking must necessarily blunder into continual bouts of

excessive money creation, and other forms of instability. To be sure, it is possible to expand credit beyond 100 percent of the gold stock, but this cannot be done for the goods and services in the economy at any given time. The attempt to do so is like trying to push down the water level in the bathtub: some of the water necessarily seeps out.

One-Hundred-Percent Money

Hayek's allegiance to the present fiduciary system is evident when he states that in holding deposits stable, banks would be reduced to "the role of brokers, trading in savings" (1933, p. 190). Rothbard (1991) offers in a slightly different context what is, in effect, a blunt rebuff to Hayek's support of the banks. In speaking hypothetically concerning the possibility of 100-percent reserve requirements, he argues that savings and deposit institutions could remain profitably in business simply by charging their customers for their services, for if they provide a useful product they would be paid for it just as consumers pay for traveler's checks. Rothbard adds:

If they [the public] are not willing to pay the costs of the banking business as they pay the costs of other industries useful to them, then that would demonstrate the advantages of banking to have been highly overrated . . . there is no reason why banking should not take its chance in the free market with every other industry. (1991, p. 27)

Hayek labels the concept of 100-percent reserve requirements as utopian in that not only will our economic progress made stagnate because of them, but bank money and notes would be eliminated and all deposits would remain fettered in savings accounts. Rothbard contends that with the elimination of fractional reserves, there will be a drop in the money supply, thus shortening credit, but that the banking industry would adjust and hold debentures of various lengths to offer as credit instead of demand deposits (1991, p. 23). In this manner, credit is potentially extended only to those who are deemed worthy of it at the market rate of interest. Since loans extended at an artificially low rate lead to an inevitable disruption of the growth process, the elimination¹⁴ of fluctuations requires the abolition of this practice.

¹⁴More precisely, we should rather speak of a radical reduction, in the sense that government would no longer destabilize the market process in this manner.

As Rothbard suggests, the banking industry as we know it would be altered dramatically in the event 100-percent reserves were required. However, its ability to grant credit will not be entirely curtailed. Only demand deposits, not time deposits, will be subjected to such requirements. More importantly, many banks have diversified into other markets such as corporate finance and various sales and trading functions. In fact, under the proposed system, the trading of debenture packaged securities could become quite profitable, similar to the field of mortgage-backed securities today.

In Rothbard's view of fractional reserve banking (1991, p. 21):

issuing promises to pay on demand in excess of the amount on hand is simply fraud, and should be so considered by the legal system. For this means that a bank issues "fake" warehouse receipts—warehouse receipts, for example, for ounces of gold that do not actually exist in the vaults. This is legalized counterfeiting; this is the creation of money without the necessity for production, to compete for resources against those who have produced. . . . I believe that fractional reserve banking is disastrous both for the morality and for the fundamental bases and institutions of the market economy.

An objection that has been used against this perspective cites the "fractional reserve parking lot."¹⁵ Here, an entrepreneur sells not the right to a parking space, as occurs in the ordinary situation, but only the right to a parking space subject to the condition that there is room in the lot for an additional automobile. The firm, then, is selling not a parking space, but in effect a lottery ticket for a parking space, where the probability of a "win" is the number of actual spaces on the premises divided by the number of such "rights" sold to the public. For example, if there are 100 parking stalls available, and the garage has sold 400 tickets, then, *ceteris paribus*, the buyer has a 25-percent chance of being accommodated when he wishes to avail himself of this service.

Now this sort of commercial arrangement, if it is conducted in an open and honest manner, is not fraudulent. It should therefore be legal. However, there is a disanalogy between this

¹⁵It would be nice to be able to cite a published claim to this effect. Unfortunately, to the knowledge of the present authors, such arguments only exist so far in the "oral tradition."

scheme and the fractional reserve system for money as currently practiced. At present, money placed in a bank is called a “demand” deposit, logically implying that it would be available, in full, whenever demanded, with a probability of certainty. If the “fractional reserve parking lot” were to be an accurate analogy to monetary practice, instead of being called a “demand” deposit, it should be called “purchasing a lottery ticket for money,” or some such. Further, in every other way—publicity, explicit contracts, etc.—banking procedures would have to be brought into line with parking lot practice. Then, and only then, could the charge of fraud be dropped. Under such conditions, there would still be the empirical question of whether or not anyone would purchase a “lottery ticket money deposit.”

This discussion should by now have made it clear that we are now very far removed from the system defended by Hayek. Yes, under certain hypothetical and narrowly stipulated conditions, something vaguely resembling the fractional reserve system defended by Hayek could be constructed so as to avoid the charge of fraud. It is certainly logically possible that someone, somewhere, might actually purchase such a ticket. But these implausible scenarios can by no means serve to justify the Hayekian analysis.

Like death and taxes, the business cycle has become invested with inevitability. With the advent of inflationary recession, something inconceivable under the Keynesian dispensation, the leaders of the economics profession are no longer so confident they can flatten out the peaks and the troughs.¹⁶ In our view, however, the former level of optimism is (potentially) justified, at least under the (admittedly politically unrealistic) assumption that government no longer generates the cycle through its destabilizing monetary policy. Under these conditions private malinvestment would undoubtedly occur, but it would result from poor entrepreneurial judgment, not centrally driven excess credit. Nor is there any reason to assume that these errors would “cluster” (Rothbard 1962), magnifying the errors of a few individuals. On the contrary, misallocation of funds, on the free market, would be dealt with in the same manner as all entrepreneurial error: with bankruptcy. But it is only with the initiation of 100-percent

¹⁶In the late 1960s, when the Keynesians were riding high, there was brave talk about ironing out the business cycle. Samuelson (1970, p. 330) even went so far as to title his chapter on the subject “Fiscal Policy and Full Employment without Inflation.”

reserve requirements, and the overall separation of state and monetary institutions, that there is any hope of stamping out the business cycle.

Rothbard contends that:

someone must propagate the truth in society, as opposed to what is politically expedient. If scholars and intellectuals fail to do so . . . all hope of social progress would then be gone, for no new ideas would ever be advanced nor effort expended to convince others of their validity. (1991, p. 43)

Hayek's initial (1931) efforts to clarify the causal connections of the business cycle were exemplary. But as we have seen, his (1933) publication—also part of Hayek I—was highly problematic. Here, then, is another bit of evidence showing not only the superiority of the Misesian over the Hayekian vision, but also indicating that although the Misesian Hayek I is preferable to the Popperian Hayek II, the former was by no means without fault.

References

- Block, Walter, and Michael Walker. 1988. "Entropy in the Canadian Economics Profession: Sampling Consensus on the Major Issues." *Canadian Public Policy* 14, no. 2 (June): 137–50.
- Böhm-Bawerk, Eugen. [1884] 1959. *Capital and Interest*. George D. Hunke and Hans F. Sennholz, trans. South Holland, Ill.: Libertarian Press.
- Buchanan, James M., Robert D. Tollison, and Gordon Tullock, eds. 1980. *Toward a Theory of the Rent-Seeking Society*. College Station: Texas A&M University.
- Buchanan, James M., and Gordon Tullock. 1971. *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Ann Arbor: University of Michigan.
- Frey, Bruno S., Werner W. Pommerehne, Friedrich Schneider, and Guy Gilbert. 1984. "Consensus and Dissension Among Economists: An Empirical Inquiry." *American Economic Review* 74, no. 5 (December): 986–94.
- Friedman, Milton. 1962. *Capitalism and Freedom*. Chicago: University of Chicago Press.
- . 1991. Friedman, Milton. "Say 'No' to Intolerance." *Liberty* 4, no. 6 (July): 17–20.
- Friedman, Milton and Anna J. Schwartz. 1963. *A Monetary History of the U. S., 1867–1960*. New York: National Bureau of Economic Research.

- Hamowy, Ronald. 1984. *Canadian Medicine: A Study in Restricted Entry*. Vancouver, British Columbia: Fraser Institute.
- Hayek, Friedrich A. 1978. "The New Confusion About Planning." In *New Studies in Philosophy, Politics, Economics, and the History of Ideas*. Chicago: University of Chicago Press. Pp. 232–46.
- . 1973. *Law, Legislation and Liberty*. Chicago: University of Chicago Press.
- . 1976. "The Non Sequitur of the 'Dependence Effect'." In *Studies in Philosophy, Politics and Economics*. New York: Simon and Schuster.
- . 1960. *The Constitution of Liberty*. Chicago: Henry Regnery. Pp 397–411.
- , ed. 1954. *Capitalism and the Historians*. [Essays by T. S. Ashton, L. M. Hacker, W. H. Hutt, and B. de Jouvenel.] Chicago: University of Chicago Press.
- . 1948. *Individualism and Economic Order*. Chicago: University of Chicago Press.
- . 1944. *The Road To Serfdom*. Chicago: University of Chicago Press.
- . 1937. "Economics and Knowledge." *Economica* 4: 33–54.
- . [1933] 1966. *Monetary Theory and the Trade Cycle*. New York: Augustus M. Kelley.
- . 1931. *Prices and Production*. London: Routledge.
- Hazlitt, Henry. 1979. *Economics in One Lesson*. New York: Arlington.
- Hoppe, Hans-Hermann. 1993. *The Economics and Ethics of Private Property: Studies in Political Economy and Philosophy*. Boston: Kluwer .
- . 1992. "The Misesian Case Against Keynes." In *Dissent on Keynes: A Critical Appraisal of Keynesian Economics*. Mark Skousen, ed. New York: Praeger.
- Hutchison, T. W. 1981. *The Politics and Philosophy of Economics: Marxians, Keynesians and Austrians*. New York City: New York University Press.
- Keynes, John Maynard. 1936. *General Theory of Employment, Interest and Money*. New York: Harcourt, Brace.
- Kirzner, Israel M. 1979. *Perception, Opportunity and Profit*. Chicago: University of Chicago Press.
- Kolko, Gabriel. 1963. *Triumph of Conservatism*. Chicago: Quadrangle Books.
- Menger, Carl. 1950. *Principles of Economics*. James Dingwall and Bert F. Hoselitz, trans. Glencoe, Ill.: Free Press.
- Mises, Ludwig von. [1969] 1981. *Socialism*. Indianapolis: Liberty Fund.
- . [1949] [1963] 1966. *Human Action*. Chicago: Henry Regnery.
- . 1957. *Theory and History*. New Haven: Yale University Press.
- . [1912] 1971. *The Theory of Money and Credit*. New York: Foundation for Economic Education.
- Paul, Ron, and Lewis Lehrman. 1982. *The Case for Gold*. Washington D.C.: Cato Institute.

- Rothbard, Murray, N. 1994. *The Case Against the Fed*. Auburn, Ala.: Ludwig von Mises Institute.
- . 1991. *The Case for a 100 Percent Gold Dollar*. Auburn, Ala.: Ludwig von Mises Institute.
- . 1982. *The Ethics of Liberty*. Atlantic Highlands, N. J.: Humanities Press.
- . 1975. *America's Great Depression*. Kansas City: Sheed and Ward.
- . 1962. *Man, Economy and State*. Los Angeles: Nash.
- . 1977. "Toward a Reconstruction of Utility and Welfare Economics." San Francisco: Center for Libertarian Studies, Occasional Paper #3.
- Salerno, Joseph. 1993. "Mises and Hayek Dehomogenized." *Review of Austrian Economics* 6, no. 2: 113–46.
- Samuelson, Paul A. 1970. *Economics*. New York: McGraw Hill. 8th ed. P. 193.
- Selgin, George. 1988. "Praxeology and Understanding: An Analysis of the Controversy in Austrian Economics." *Review of Austrian Economics* 2: 19–58.
- Smith, Adam. [1776] 1965. *An Inquiry into the Nature and Causes of the Wealth of Nations*. New York: Modern Library.
- Wieser, Friedrich von. 1967. *Social Economics*. A. Ford Hinrich, trans. New York: Augustus M. Kelley.
- Williams, Walter, E. 1982. *The State Against Blacks*. New York: McGraw-Hill.

The Myth of the Income Effect

Pascal Salin

Microeconomics traditionally splits the consequences of price changes into two different effects: the substitution effect and the income effect.¹ The effects are supposed to work in diverging directions. The precise reaction of demand or supply to a relative price change can operate either way. The supply curve is not necessarily upward-sloping and the demand curve is not necessarily downward-sloping.

The income effect plays a crucial role in mainstream microeconomics, as well as in applied economics, because it induces economists to think, for instance, that there are backward bending curves. Such curves are particularly assumed to exist for the supply of labor. Within a certain range, a decrease in the real wage would not bring about a decrease in the supply of labor, since people want to maintain their income and, therefore, increase their supply of labor. If this is true in tax theory, the “Laffer-effect”² would not exist, at least in certain conditions in which the supply curve for labor is supposed to be atypical. With the income effect, an increase in the tax rate on labor income—i.e., a decrease in the real after-tax wage—would be compensated for by working more. The “Laffer-effect” assumes that there is always an inverse relation between the tax rate and productive efforts. In a similar way, the income effect would

Pascal Salin is professor of economics at the Université Paris-Dauphine.

¹I thank Philippe Lacoude and Frédéric Sautet for their very pertinent comments on a previous version of this paper, the idea of which actually stemmed from a talk with them.

²The increase in a tax rate decreases the tax base and, beyond a certain tax rate, this effect is sufficiently important for the increase in the tax rate to induce a decrease in tax revenues instead of an increase. It is now usual to label this process as the “Laffer-effect.” But it had been developed before Arthur Laffer gave it its popular expression. One can find it, for instance, in a refined form, in Geoffrey Brennan and James Buchanan, *The Power to Tax* (Cambridge University Press, 1980).

The Review of Austrian Economics Vol. 9, No. 1 (1996): 95–106
ISSN 0889-3047

account for a possible increase in the rate of savings when the return on savings is decreasing.

The existence of the income effect assumes that the concept of income can be defined in a non-ambiguous way. It can be measured and quantified. In fact, this is not true. Income can only be understood using a precise concept of utility which does not allow room for measurement. In the present paper we show that the income effect is not consistent with a purely logical theory of utility. In reality the income effect can exist only if one adds a specific assumption. Contrary to the belief that the income effect is a general principle, it is in fact only a possible consequence of a specific assumption. The substitution effect—consistent with the general theory of utility—is the only general principle.

Let us consider an individual's supply curve for labor in a world in which there are only two goods, leisure and wheat. Time is the only scarce production factor. Land is not operationally scarce, so that the physical productivity of one hour of labor in terms of wheat is constant. There is a certain relative price between leisure and labor (the production of wheat). In one hour, one can produce one hour of leisure or one pound of wheat. For this given relative price, an individual wants to share his available time between leisure and the production of wheat. If he does not desire to produce and to consume more wheat, it is because sacrificing one hour of leisure to get one more pound of wheat is not worthwhile from his own point of view.

Now, economists usually define the income of an individual by the quantity of tradable resources (wheat) he is producing at each period; his production of leisure services is excluded. But such a notion of income is arbitrary in comparison with the more general concept of satisfaction (or utility): The individual's total satisfaction is equal to the "sum" of the satisfaction he draws from leisure and the satisfaction he draws from consuming the wheat he has produced. Saying that the desired level of income is equal to y is in fact saying that a maximum satisfaction is obtained when the production of wheat—named income—is equal to y and the production of leisure services is equal to x , knowing that $x + P_x y = t$, where P_x is the price of one unit of wheat in terms of hours of leisure and t is the available quantity of time, for instance 24 hours a day.

Let us assume now that a disease strikes the wheat crop and it cuts the productivity of labor by one half or—the equivalent—a tax has been introduced on the production of wheat, so that one

can keep only half a pound of wheat for one hour of work. The opportunity cost of producing and consuming wheat increases in terms of leisure. The logic of choice would imply that one does not produce as much wheat as before.

Referring to this logic of choice, what could be the meaning of the income-effect? It implies that the individual aims at maintaining the level of his income, at least partially. That is, he wants to maintain his production and consumption of wheat. But it is inconsistent to hold that the individual has two aims in mind. He would have an income target (a certain amount of y): leisure time would be nothing but an instrument which can be used to reach this income target through a process of transformation. On the other hand—and contradictorily—he would have a satisfaction target: he aims at maximizing his level of satisfaction by obtaining an optimal mix of x and y .

In the above example, it is traditional to say that the individual facing an increase in the cost of wheat may not decrease his production of wheat—although it is most costly in terms of sacrificing leisure—but may decide to work more. Therefore, the supply curve would be backward-sloping.³ There are more work hours when the return of one hour of work is decreasing. The income effect would thus dominate the substitution effect. In fact, the belief in the existence of the income effect implicitly requires *specific assumptions* which are not absolutely and logically necessary. This is to say, the income effect does not exist as a general phenomenon; it is a mathematical illusion in a badly specified world.

What happens in reality? If one assumes that an individual increases his supply of labor when the return on work hours decreases, it is necessarily because one has added an additional assumption (if not, one ought to admit that the substitution effect—which necessarily exists—is the only one to be met). Very likely, it is implicitly assumed that wheat—contrary to leisure—is necessary for one's survival, so that, if the substitution effect was the only existing effect, there would be a risk for an individual of passing under the subsistence level, which he cannot accept. But that is really saying that there are three goods (or three sources of utility) and not two, namely the pleasure

³Hans-Hermann Hoppe raised an interesting question on a previous version of this paper: If the income effect does not exist, does it imply that a supply curve cannot be backward sloping? We explain in the appendix why, in our opinion, it is impossible.

drawn from leisure, the pleasure drawn from consuming wheat, and survival services obtained from wheat consumption. One also assumes implicitly that there is only one possible technology to get survival services, namely wheat production (which is called "income"). But why would there not be another one? As an example, it may happen that the concerned individual knows that he can survive by extracting roots from the soil, the taste of which he does not like very much, but the nutritive value of which is indisputable. So long as wheat is not expensive (in terms of time) in comparison with roots, he prefers to eat wheat. If the opportunity cost of wheat production increases too much, he will certainly consider possible substitutions between wheat, roots, and leisure.

It may happen that shifting to another consumption pattern cannot be done immediately, since one has to find the most productive field of roots, or to invent suitable instruments to extract the roots. Therefore, during some period of time, the individual may increase his production of wheat in order to survive and to prepare the change in technology. But this is also saying that the existence of the income effect is due to two specific assumptions: The existence of some subsistence level and the existence of transfer costs when shifting from one activity to another one. Now, if the individual could have forecast the relative price change between wheat and leisure, he would certainly not have waited until the last moment, which obliges him to increase his supply of work in order to produce wheat. He would probably have decreased his leisure time earlier in order to prepare the new survival technology. Because of the increasing price of wheat relative to roots, there is a substitution effect between both these goods which make it possible to meet in different proportions both targets (survival and gastronomic pleasure).

The apparent appearance of an income effect is due to a model which has been badly specified and surreptitiously adds assumptions which can be interpreted as the introduction of an exogenous event in a world in which there are rigidities of a technical or institutional nature. Choices are then reshuffled and, during a transition period, what is named the income effect may appear in certain circumstances, as the result of a search for minimizing adjustment costs until the process of change in activities has led to all desired substitutions.

The above remarks can be reconciled with traditional statements of microeconomics. Let us take the traditional figure which is used to illustrate the existence of both the substitution and the

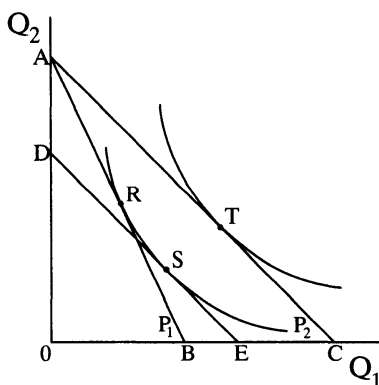


Figure 1

Source: James M. Henderson and Richard E. Quandt, *Microeconomic Theory* (New York: McGraw-Hill, 1958), p. 26.

income effects (figure 1).⁴ The initial equilibrium position of a “consumer”⁵—who has to choose between two goods, Q_1 and Q_2 , for a relative price P_1 —is point R. If the price of Q_1 decreases, the new budget constraint is AC and the equilibrium point becomes T. Now, one can obtain this later point by a formal splitting of the path from R to T into two paths: The path from R to S represents the substitution effect (pure relative price change, the utility level being maintained constant), and the shift from S to T represents the income effect (the decrease in the price of Q_1 corresponding to a gain in the purchasing power of income). Such a presentation does not help to clarify an important aspect of the problem, namely that the rotation of the budget constraint from AB to AC implicitly assumes that the consumer holds a stock of goods Q_2 equal to OA, so that the relative price change between Q_1 and Q_2 does not affect the position of point A: For a stock of goods Q_2 equal to OA, the individual can obtain a different (higher) quantity of Q_1 . On the contrary, if the individual was holding (or producing) good Q_1 , it would be point B which would

⁴Figure 1 is extracted from the classic textbook by James M. Henderson and Richard E. Quandt, *Microeconomic Theory*. But a similar graph can be found in most microeconomics textbooks.

⁵Let us use this traditional concept, although it induces the danger of splitting a human being into different, non-coordinated, parts (the consumer, the producer, the taxpayer, etc.).

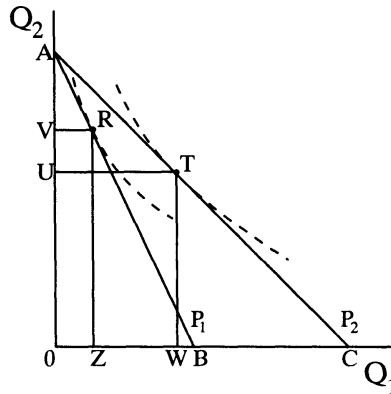


Figure 2

stay at the same place and the budget constraint would then cut the Q_2 axis at a point lower than D. Total utility would thus diminish. The result is perfectly logical: If the individual supplies Q_2 and demands Q_1 , a decrease in the relative price of Q_1 (increase in the relative price of Q_2) leads to the transformation of more Q_2 in goods Q_1 . The graph can then be interpreted from three different points of view.

First Interpretation

The relative price change reflects a technological change. Thus, with the same quantity of work, OA, an individual can obtain a greater quantity of wheat (OC instead of OB). The figure then does not represent the choices of the abstract "consumer" of textbooks, but the choices of the individual who *acts* (for instance Robinson alone on his island). If the productivity of labor is increasing, the individual will produce more wheat. On figure 2, OA represents the number of available hours per day, namely 24 hours. When the relative price is P_1 , the amount of work is equal to AV, the quantity of leisure to OV and the quantity of wheat produced to OZ. When the price of wheat in terms of work hours shifts to P_2 , work increases (AU), leisure decreases (OU) and the quantity of wheat produced increases (OW). The marginal sacrifice which is necessary to obtain one more unit of wheat having diminished, one substitutes wheat to leisure. There is a pure substitution effect.

Second Interpretation

Figure 2 now represents an exchange economy. The individual can obtain wheat (Q_1) by selling tomatoes (Q_2) which he is producing. Given the existing techniques and the limitation of his available time, the maximum quantity of tomatoes he can produce is OA. If the wheat price decreases from P_1 to P_2 (the price of tomatoes increases), he will sell more tomatoes to get wheat: On figure 2 his sales of tomatoes increase from AV to AU and his consumption of his own tomatoes decreases from OV to OU. His income does not change if it is measured in terms of tomatoes; it increases in terms of wheat. But there is only a process of substitution between both commodities and no income effect.

Third Interpretation

Figure 1 could be interpreted as representing the *partial* behavior of an individual *only* in his role as a consumer. He is a producer of another commodity, Q_3 , which is not represented on the figure, and he demands Q_1 and Q_2 . When the budget constraint AB is rotating to the position AC, one implicitly assumes that the relative price decrease of Q_1 in terms of Q_2 does not affect that valuation of Q_3 —the supplied good—in terms of Q_2 : The supply of commodity Q_3 in exchange for Q_1 and Q_2 would remain equal to OA (in terms of Q_2), which is debatable. In fact, there is a decrease in the price of Q_1 relatively to Q_3 , which ought to affect the supply of Q_3 . The new budget constraint is very likely situated between AC and DE (figure 1).⁶

These various interpretations are useful since they make it possible to clarify what is precisely the behavior of the individual who acts. But they may also help to demonstrate why the mainstream distinction between the substitution effect and the income effect is misleading. From a purely formal point of view, one can distinguish, as we have seen, the shift from R to S and the shift from S to T (figure 1). But these shifts do not correspond to any *actual* human action.

Let us take the first or the second interpretation.⁷ It is meaningless to say that point S corresponds to a pure substitution

⁶As an example, one may imagine that the budget constraint is rotating around point R.

⁷Both interpretations are very close to one another. They imply that, from a given stock of resources—for instance time resources—an individual can obtain a good either via a technological process of transformation he is doing by himself, or via a process of exchange involving his labor services or the outcome of his labor.

effect (with a constant level of utility), since point D has no meaning for the individual. In fact, he *has* resources equal to OA and there is only a substitution effect, expressed by the rotation of the AB line to AC: Given the time resources he has, substitutions are possible along the budget constraint and these substitutions depend on the relative price between Q_1 and Q_2 .

In the third interpretation, one cannot determine the value of the so-called income effect as long as one has not made clear what are the actual substitutions between the quantity of good Q_3 , on the one hand, and the quantities of goods Q_1 and Q_2 on the other hand. But this is precisely saying that there is no income effect, and that there are only substitution effects between several possible uses of resources. Let us take the case of an individual who is choosing, at some point of time, a given allocation of his resources, specially as regards his time resources, to obtain a certain basket of goods, among which leisure has to be considered. To accept sacrificing an additional unit of leisure, he necessarily has to get compensation by obtaining a greater quantity of other goods. If ever there is an increase in the market price of time (a decrease in the price of other goods), it creates purely and simply a substitution effect.

The general belief in the existence of an income effect partially stems from an artificial separation between consumption activities and production activities. This distinction is usual in microeconomics: The theory of consumption and the theory of production are founded on different grounds and different behavioral assumptions. In fact one does not supply a good without demanding another good at the same time, so that the decision to produce and to supply cannot be separated from the decision to demand and to consume. By cutting, for instance, decisions related to consumption or decisions related to work from other decisions, the traditional theory implicitly maintains as constant certain variables which cannot be maintained as constant. It considers, for instance, the choice between two consumption goods, as if the whole set of activities was not implied. A specific assumption concerning choices made about these other activities is then called the income effect. As we have seen, a change in the relative price between two commodities has consequences which depend not only on the choices made by an individual about these commodities, but about the whole set of possibilities open to him.

In mainstream microeconomics, the income effect and the substitution effect are presented as similarly general effects. In fact, the substitution effect is the only one that actually exists. It

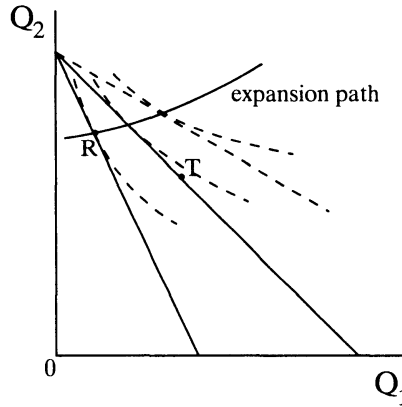
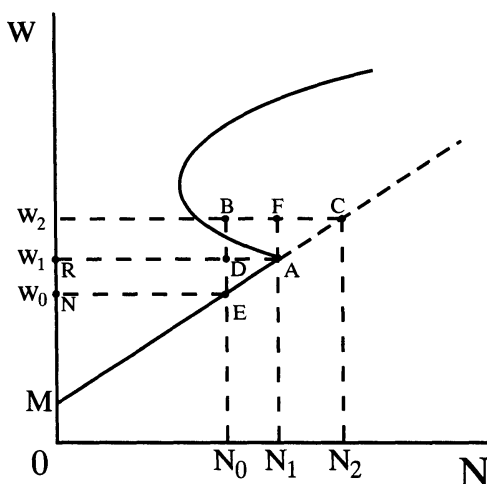


Figure 3

may appear that something close to the so-called income effect appears under specific assumptions when one adjusts from one equilibrium point to another one. However, this effect, which describes a specific adjustment path, is only transitory and it does not necessarily exist. This difference between both effects has important consequences. It means, for instance, that the Laffer effect necessarily exists: An individual shifts from an activity which is more heavily taxed to other activities. However, the process of adjustment may not be instantaneous, so that one may have the feeling that people react contrary to what the substitution effect would lead them to expect. However, this effect, if it ever exists, is necessarily a transitory and secondary phenomenon.

From a formal point of view, one can draw an expansion path so that the quantity of labor is decreasing as long as its price is increasing (figure 3), but this is nothing else than a graphical representation. Now, a graphical assumption does not give any information on the real motives of human behavior. The latter has to be explained by the law of decreasing marginal utility which is in reality a theory of human choices founded on rational logic. If one believes—as we should—that all reasoning in the social sciences has to be coherent with rational logic, it just means that one must get rid of the graphical or mathematical appearances which are inconsistent with this logic.

**Figure 4**

Appendix

Is a Backward-Bending Supply Curve Possible?

As we mentioned in footnote 3, an interesting question can be raised: Is it consistent to admit that both the income effect does not exist and that, however, the supply curve (for instance for labor) can be locally backward-sloping? To try to solve this question, let us take the case of an individual who is going up his supply curve when his real wage increases (figure 4). His supply of labor then shifts from E to A when his real wage shifts from w_0 to w_1 . This increase in supply obviously corresponds to the law of diminishing marginal utility: In order to give up one additional unit of leisure, the individual wants a higher real wage as a compensation for his loss of utility. This is exactly the meaning of the supply curve.

However, one has to clarify the distinction between the marginal return and the average return. Let us assume, as an example, that the supply curve on figure 4 represents the daily supply of labor of the individual. At point A, he works during 10 hours for an hourly wage of \$1, so that his daily return is \$10. He considers that it is worthwhile working one more hour only if this additional hour is paid at least \$2, which means that he wants at least an income of \$12 for 11 hours of work. From his point of view,

it is exactly equivalent to get \$2 for the marginal hour, the other hours being paid \$1, or to get a higher average hourly wage—in our example \$1.09—under the condition that he works 11 hours and not 10. If his job is totally specific—there is no possible substitution with other works—the price of successive hours can be differentiated and the employer may be ready to pay more for the additional hour.⁸

Now, if we consider a market situation with substitutable workers, such a differentiation of wages according to the number of hours worked within a day is not possible, so that the supply curve of figure 4 represents the amount of work hours offered for any hourly (average) real wage. Let us anyhow consider the behavior of one given worker. Let us assume that he is at point E when the real wage is equal to w_0 . The precise meaning of this point is the following: Any wage lower than w_0 would be insufficient to compensate the individual for the disutility of an amount of labor equal to N_0 . He needs at least w_0 or more. However, if the market wage is w_0 , he cannot get more than w_0 . Now, if the market wage rate increases from w_0 to w_1 , the preferred point is A. As there is no wage differentiation, the new wage is paid not only on the marginal hours ($N_1 - N_0$), but even on those hours which were already worked when the real wage was w_0 , namely N_0 . The change from w_0 to w_1 thus represents the minimum increase in the *average* wage rate which is necessary for the *marginal* return thus obtained to be considered as a sufficient compensation for the increased disutility of work.

Under the traditional neo-classical approach it would be said that the individual got a surplus equal to MEN and that the increase in the market wage brought him an additional surplus equal to NEAR. In fact, the individual was ready to offer N_0 for a wage equal to w_0 . Now, he can obtain w_1 for this amount of work, so that he gets a sort of windfall profit equal to DENR; and he also gets a marginal surplus equal to EDA. For reasons linked with the distinction made above between the average and the marginal wage rate, the existence of this surplus can be disputed: One might say that the increase in the average wage rate ($w_1 - w_0$) is just sufficient to allow the individual to get the minimum marginal wage rate he wants in order to offer additional work

⁸There is always a certain degree of specificity in the employment of a worker. For instance, as transaction costs do exist, an employer will prefer paying more for extra hours by a worker who is already employed in his firm than hiring an additional worker.

time (from N_0 to N_1). Therefore, we cannot quantify something like a surplus. But we can be certain that point A is preferred to point E. For a wage rate equal to w_0 the individual did not want to work more than N_0 . An increase in the average wage rate necessarily means an increase in the marginal rate, thus inducing him to work more: He prefers A to D. This is the rationale for the upward-sloping supply curve.

This discussion on the very meaning of the supply curve may help us to decide whether this curve can become backward-sloping at some point, for instance at point A. If such a situation exists, it means that, if ever the wage increases from w_1 to w_2 , the supply of labor decreases from N_1 to N_0 (instead of increasing to N_2). This is exactly what is assumed by those who believe in the existence of an income effect: As the individual gets more money for any work hour, he works less hours. Is that possible?

Saying that the curve is backward sloping, as represented on figure 4, is saying that point B is preferred to A.⁹ As we do know, point B is preferred to point D and to point E (the higher is the wage rate, for a given amount of work, the more satisfied is the individual). We also know that D was not preferred to A, but there is no reason for B not to be preferred to A. However, the real question is not knowing whether B is preferred to A, but knowing whether it makes sense saying that B is on the supply curve. We explained above why A is preferred to D (or, at least, not less desired). Exactly the same reasoning has to be made as regards point B: If the wage rate increases from w_1 to w_2 , the individual desires a point which is not on the left of point F (he prefers for instance point C). Therefore, although it may be true that B is preferred to A, C is preferred to B. It means that point B is not a point that the individual may *actually* choose.

Another way to look at this problem is as follows. As we saw above, the supply curve is nothing but a frontier between points which are not desired (those under the curve) and those which are accepted (those which are on or above the curve). At point E the individual declines all wage rates lower than w_0 and accepts all higher wage rates. At point B, the individual is supposed to decline all wage rates lower than w_2 and to accept all higher wage rates. But it does not make sense to assume both that wage rates higher than w_0 are accepted for an amount of work equal to N_0 , and that wage rates under w_2 are declined for the same amount of work.

⁹If B was not preferred to A, it would just mean that the supply curve cannot go through B.

Hayek: Some Missing Pieces

Anthony de Jasay

Has Hayek a theory of the social order, a comprehensive view of society as a determinate and self-reproducing system, in the same way as Hobbes or Marx, though perhaps no one else can be said to have an (albeit simpler) theory of the social order? The question has some weight, since if it is true that Hayek is the most influential twentieth-century advocate of liberal government, the strength and penetration of normative liberal doctrine must to no small extent depend on the persuasive force of his positive social theory: the less coherent it is, the more modern liberalism is vulnerable to erosion and invasion by incompatible elements.

It is hard, probably harder than it looks at first sight, to be sure what we mean by a social order. Hayek himself mostly uses the term “order” in the context of his ideas about the spontaneous order. However, he intends order as such to mean something akin to a pattern or schema, such that by looking at a part or a phase of it, we can make good enough conjectures about the whole (1973, p. 36). It is as if, by finding a piece of a jigsaw puzzle that depicts a cloven hoof, we could tell that the puzzle, if it were all fitted together, would in all likelihood represent a cow, a goat, or perhaps the devil, but certainly not a lady with her parasol.

My thesis, putting it at its sharpest, is that Hayek shows us pieces of a complex jigsaw that are intriguing and inspiring, but they do not suffice to let us predict whether, if we had all the pieces, the completed puzzle would show a cow, a goat, or the devil. If I am anywhere near right, he has no complete theory of the social order to back up his liberal recommendations. They are in any case a little incongruous, since he predicts that whichever kind of order is superior will duly prevail through group selection,

Anthony de Jasay is an independent scholar now living in France.

The Review of Austrian Economics Vol. 9, No. 1 (1996): 107–18
ISSN 0889-3047

hence recommending it is somewhat out of place even if we could know that it was not superior, and *a fortiori* if we could not. Does Hayek show us how to tell?

There are, it seems to me, at least three areas of the puzzle where he has left out important pieces, and where pieces he did place do not really fit together. One is the distribution of the social product, the other is public goods, and the third is the spontaneity of the very spontaneous order that gives the whole puzzle its character.

Making the Free Society Attractive

One of Hayek's most widely known normative ideas is that there can be no such thing as social, or distributive, justice. The concept is simply a category mistake: "there can be no distributive justice where no one distributes" (1978, p. 58). It is certainly doubtful whether one can defend, from arguments drawn from justice, the popular belief that certain participants in arm's length transactions are responsible for the distributive shares that accrue to other participants; that rich employers should pay higher wages to poor employees, and that supermarkets should not drive small shopkeepers out of business.

Hayek holds, reasonably enough, that the terms of voluntary exchanges are determined objectively, they are not matters of anybody's good intentions, nor of what richer parties think the poorer parties ought to be getting: nobody distributes. What he calls the "market order" entails a distribution that is neither just nor unjust. It is, however, efficient. As such, it has instrumental value and can serve other valuable ends. These ends, for Hayek, cluster around the maximum chance for the randomly chosen individual to conduct his own life successfully. Distributive justice does not figure among the ends the riches created by the free market order should serve. A long line of others, from John Stuart Mill to A. Mueller-Armack and beyond, have of course taken the well-known position that the market order conforms to "economic laws" and is neither just nor unjust, but the wealth its efficiency creates can be devoted by "society" to satisfying, among other things, the requirements of distributive justice.

Both Hayek's and Mill's position, different as they are from each other, stand in sharp contrast to the more rigorous thesis that the market order does in fact produce a distribution that *is* just if the distribution of ownership in a selected initial position was just, and subsequent exchanges were free from force and fraud.

Hayek, in his insistence that the whole issue is categorically irrelevant to the market order, leaves a blank where others put distributive justice. Its missing piece is an unintended invitation for them to fill the void.

However, while disputing the very sense of the concept of distributive justice, Hayek is nevertheless concerned with distribution. He notes that modern societies without exception do organize welfare states,—is this cultural selection at work?—and that this evolution is consistent with respect for “abstract rules of just conduct.” It is more than a historical shift in the spontaneous order, more than an unintended result of human actions directed at other purposes: deliberately helping it along is a positive “task of the defenders of liberty” (1960, p. 259). “[T]hough a few theorists have demanded that the activities of government should be limited to the maintenance of law and order, such a stand cannot be justified by the principle of liberty” (*ibid.*, p. 257). The welfare state is a conglomerate of many diverse elements, some of which may “make a free society more attractive” (*ibid.*, p. 259). For one, government must provide for “the minimum of sustenance” for the helpless, and this minimum should be, not absolute, but relative and rising with the general standard of living (*ibid.*, p. 285). For another, such provision cannot be confined to the deserving poor, but must be extended to all (*ibid.*). As a corollary, it becomes the recognized duty of the public to compel all to insure or otherwise provide against the “common hazards of life” (*ibid.*, p. 286).

It is clear enough that a guaranteed minimum income, once granted, will not for long be kept down to the level of absolute physical subsistence, but will creep upwards and take on features of a defense against “relative deprivation.” It is also clear that if people no longer have an incentive to provide against bad luck and old age, a case is created for compulsory social insurance. What is less clear is why Hayek considers, not only that these things are bound to happen, and to happen on a scale that expands with economic progress, but that it is no bad thing that they should. Their object is not to conform to any moral imperative, whether of compassion, fellow-feeling, let alone distributive justice, which he rejects as the product of muddled thought. Nor is it because a distribution brought about by the market order can, along the lines of Benthamite and Pigouvian utilitarian thought, be “corrected” to generate a larger sum of aggregate utility, for Hayek to his great credit never embraces the idea of interpersonal aggregation of utility. In fact, his advocacy of such

quintessentially redistributive measures as the guaranteed minimum income, is accompanied by injunctions that such measures must not have a redistributive purpose! Barring both justice and utility as their object, all he leaves us with by way of justification is that some redistribution, (if it is not intended to be redistributive) “makes a free society more attractive.”

Why, however, should one seek to *make* the free society more attractive? Is it not going to prove itself more attractive anyway in cultural selection, by its superior aptitude to prevail over less free societies (assuming that the ambiguities of what it means to “prevail” have been resolved)? Manifestly, Hayek thinks it can do with a bit of help. Embellished by the institutions of a moderate welfare state, presumably more people will opt (vote) for it, or fewer people will desert it for the totalitarian alternatives he abhors. Does this mean, however, that the free society is not the social order that prevails because its intrinsic properties make the groups that adopt it more numerous?—does it mean that the free society is a social order that prevails over others if and because its properties are adjusted to what people from time-to-time find attractive (even if it does not make the groups adopting it more prosperous and numerous)? If the latter is the case, what distinguishes Hayek’s social theory, or at least the part he has made explicit, from the theory of democracy as the system where social choices are made by adding together votes for alternatives being acknowledged as superior, not by virtue of its intrinsic capacity to make its host group grow, but by virtue of attracting more votes?

Enough pieces are missing from the jigsaw to permit either interpretation, though Hayek would no doubt protest quite sharply against his cultural selection being, by a piece of impudent distortion, equated with procedural democracy. However, his missing pieces leave room for a “conglomerate of elements” that may well be mutually inconsistent, part cow, part goat, part devil, leaving it to the spectator to call which is which. What exactly is a free society? On what grounds can Hayek predict that it will prevail? Above all, what is the point in its functioning of the unplanned interplay of individual decisions whose collective effects are unintended, and what of consciously formed collective choices carried out by the agency of the state? Hayek’s *obiter dicta* on redistribution seems to me to leave the question largely open. This void is only deepened by his treatment of public goods and the role of the state in providing them.

The Pivot Between Ordered Anarchy and Statism

There is a measure of unself-conscious irony in Hayek's call for a "much more clear-cut attitude towards [public goods] than classical liberalism ever took" (1978, p. 144). Classical liberalism would entrust to the state the provision of only one, very special, public good, to wit, "the enforcement of general rules of just conduct" (ibid.). There is, however, a multitude of other "highly desirable" public goods that "*cannot* be provided by the market mechanism" because they "*cannot* be confined to those who are willing to pay for them" (ibid., my italics). Therefore the means for providing them are either raised by the coercive power of the state, or not raised at all. The liberal may wish that the way be left open for private enterprise to provide them if a method is discovered for it to do so (1978, p. 145), but pending such discoveries, it is legitimate and indeed mandatory for the state (or local authorities) to tax society in order to enable it to enjoy these highly desirable things.

Is Hayek's position "much more clear-cut" than that of classical liberalism? Public goods are of central importance to social and political theory. If they *cannot* be provided in voluntary transactions, but we must and want to have them, the state is necessary and Pareto-superior. If they *can*, ordered anarchy is possible, and the state usurps the space that, in its absence, would be filled by Pareto-superior, voluntary transactions. It may be held that property and contract enforcement is a necessary condition of voluntary transactions. It may be further held that such enforcement is a public good only the state *can* provide. This, in brief, is the classical liberal position of the minimal, protective state. It may or may not be good theory, but it is clear-cut enough. More recent theory suggests that even contract enforcement *can* be provided voluntarily by those who expect to benefit from respect for their contracts, and there is no evidence that organizing a state for the enforcement is more efficient, less costly in terms of total transaction costs, than its decentralized, private provision. In this view, even the minimal, classical liberal state is a needless blemish upon ordered anarchy, let alone the modern liberal state subsumed by Hayek, that has a mandate to tax society for the sake of providing goods and services as long as they are both public and desirable.

Hayek seems strangely unaware of the pivotal role of public goods theory between ordered anarchy and statism, and treats it

cursorily. Hardly realizing its consequences, he accepts the textbook division of the universe of goods and services into two exogenously determined halves, public and private. Private goods are *excludable*, hence *can*, public goods are *non-excludable*, hence *cannot* be produced in voluntary transactions, where goods are forthcoming against equivalent resources or not at all.

In reality, there is no such exogenous division. Nothing is "excludable" without further ado; for nothing can be sold without the seller incurring costs to exclude from access those who would not pay the price. Exclusion cost is no more avoidable in a good destined to be sold than is the cost of production or transport. Everything is excludable at *some* cost that may be high or low, depending on a host of circumstances, of which the physical characteristics of the good is only one. Over the universe of goods, exclusion cost is a continuous variable. Where society draws the dividing line between public and private goods is an endogenous decision, for social theory to define. Providing a good publicly saves exclusion cost. This advantage may be partly, wholly, or more than wholly offset by costs arising from wasteful use of the good the consumer can have without paying for it, and from other, less direct risks. If social choice were usually "collectively rational," goods would be provided publicly if the saving of exclusion cost outweighed the disadvantages and added costs of publicness. As it is, whether a good becomes public, or stays private, is decided by the "public" through a political process that is not set up, and is quite unlikely, to be "collectively rational" in the above sense. Certain goods become public goods because it is held that people ought not to *have* to pay to have them, others because they *won't*. All this is well understood now, and was already understood when Hayek expressed his view that the state ought to provide "highly desirable public goods."

The half-universe of public goods is in fact one *we* fill. It thus comes to contain innumerable goods that are desirable *if* and because they are public, so that their marginal cost to the individual consumer of the good is nil or imperceptible, and they amount to a "free lunch," to something for nothing. If so, the observation that they are highly desirable is a product of circular reasoning. As long as the good remains a good, (i.e., short of saturation) every potential consumer of it will readily vote for its public provision if it is not yet so provided, and for its provision on a more generous scale if it is provided but sparingly. Where should the line be drawn? How should a liberal society count the

votes for more of everything, and the votes against the taxes to pay for it? Whichever way it counts them, it has relatively little chance to stay liberal.

There is only a missing piece in Hayek's theory where the principles should be that a liberal society would adopt to draw the line between public and private, to keep it there, and stay liberal in the process. Unlike classical liberalism that confines the state to the provision of a single public good, law enforcement, Hayek's social order is less, rather than more, clear-cut: it permits, if not positively mandates, the state to produce any number in any quantity; the state's place in society is consequently *ad hoc*, open-ended, indeterminate, and no amount of dire warnings against socialism, fatal conceit, and loss of freedom will make it more determinate.

A theory of social order is incomplete if it makes no serious attempt at assessing the long-term forces that make the public sector grow or shrink. This can hardly be done without relying on a defensible theory of public goods. Hayek feels no necessity for one. Strangely, the question seems to have held no interest for him. By way of making good the missing piece, one must insert some account of the conditions under which goods will remain private, produced only for restricted access against payment in full, as opposed to the conditions that will favor their production for unrestricted access by any member of a given public, with the necessary resources being raised either by voluntary association under contract, or by involuntary taxation. The relative weight of these three alternative solutions is perhaps the decisive influence on the extent to which a society is political ("politicized" is the pejorative word usually employed for it), shaping its life by collective rather than by individual choices. The question is of abiding interest to Hayek. He does not answer it, though he fervently wishes throughout his massive *oeuvre* that the answer should favor individual choice.

Who Enforces the Enforcement?

Why does it matter to Hayek, or anybody, whether an order is spontaneous or not?

The attraction of spontaneity is both moral and prudential. Though it is not clear whether Hayek saw more than instrumental value in it, he stressed that the elements in a spontaneous order "arrange themselves" rather than being arranged by "unified direction" (1960, p. 160). When the elements are human beings, their property and their choices, nobody's dispositions are

imposed on him by another's command. Everybody chooses for himself what seems to him the best, given that everybody else chooses likewise. All choices are interdependent, and made mutually compatible by property rights and their voluntary exchanges. None dominates and none is subordinated. This lends the order in question a moral *laissez passer*, while non-spontaneous orders, constructed by imposing some alternative on the participants by authority or the threat of force, are morally handicapped by their coercive element. If they are to pass for legitimate, they need to show some compensating merit. Spontaneous social orders, in other words, have a *prima facie* moral standing. Constructed orders must first earn it, or do without.

The prudential attraction of spontaneous orders springs from the belief, strongly held by Hayek and fairly well supported by historical evidence, that since the knowledge required for successfully designing a complex order is either irretrievably dispersed or latent or both, the constructed order runs a high risk of being inefficient if not grossly counter-productive.

Game theory calls "coordination game" an interaction where, if all or most players adopt the same norm of behavior (strategy), all get a payoff that is no worse and may be better than if they adopted different norms. Compliance at least weakly dominates deviation. Hayek's spontaneous order is at first sight a coordination game: he speaks of rules that, if they are generally observed, make all members of a rule-following group "more effective," "because they give them opportunities to act within a social order" (1978, p. 7, Hayek's italics). The rules are randomly generated, by analogy with genetic mutation. Some are positively selected in a process of "cultural transmission . . . in which those modes of conduct prevail which lead to the formation of a more efficient order" (ibid., p. 9), because the more efficient order helps the group living within it to "prevail" over other groups. A classic and appropriately Austrian example is the use of money, a more efficient "norm" or "rule" than barter. No member of the money-using group can do better by reverting to barter once most others trade against money. Compliance dominates deviation.

All would be well if Hayek confined his concept of the spontaneous order to cases of voluntary rule-following that are coordination games, i.e., where the emergence of the order depends on some members of a group adopting the same rule of behavior, but once they do, the order is self-enforcing: all members have a continuing incentive to adhere to it and can only do worse for themselves if they deviate from it. Patently, however, there are

important rules that do not function like this. Once they are widely followed, they generate an incentive for the individual member of the group to violate them. Perhaps the simplest “spontaneous” order, as Hayek would call it, that operates in this perverse way is the queue. Every member of the group that has a rule of queuing rather than milling around and pushing each other, gains from every other member following the rule. However, the member who jumps the queue gains more than the one who stands in it; he can abuse the decent restraint of the others. Queue-jumping dominates queueing. The same is true of the spontaneous order that is at the center of Hayek’s theory, the “market.” It will not function to the advantage of every participant unless at least two key rules of conduct, respect for the property of others and performance of reciprocal promises, are widely followed. However, if they are followed by some, this *ipso facto* tempts others to steal, usurp, trespass, and default on contracts. These favorite deviations offer a higher “payoff” than compliance with the rules, which of course renders compliance a potentially self-destructive mode of conduct. Neither queueing, nor the market, nor many other ostensibly spontaneous orders are truly spontaneous, i.e., coordination games along the benign, self-enforcing lines of using money, speaking the same language or driving on the same side of the road. They are thinly disguised or overt prisoners’ dilemmas.

Though he steers clear of game terminology, Hayek is quite aware that this is so, and that those of his putative spontaneous orders that are in effect prisoners’ dilemmas, and have deviation as their dominant strategy, need something more than the efficiency of their rules of conduct if they are to survive. Not being self-enforcing, they need some support from rule-enforcement. At one point, he suggests that the successful group, though it does not realize to which rule it owes its superiority, “will accept only those individuals as members who observe the rules traditionally accepted by it” (1978, p. 10). Hayek’s group, then, expels robbers and cheats. It uses ostracism to punish and deter violations of its rule. Ostracism is one of the several time-honored voluntary enforcement mechanisms that have been employed, since the dawn of civilization, to ensure the survival of beneficial but fragile conventions, including adherence to the customs and laws of property and contract, where the convention itself generates an incentive to break it. Ostracism, like other defenses against violation, can thus be understood as an auxiliary convention, a satellite serving the fragile, nonself-enforcing main convention.

In the absence of such supporting conventions, the emergence and widening of the division of labor, trade, and capital accumulation would be incomprehensible. So would be those cultural, legal, and political institutions whose material wherewithal was produced by these developments.

There is some excuse for holding that a spontaneous order that needs to be enforced is still a spontaneous order, if its enforcement itself is spontaneous, the norm of a voluntary convention. Its adherents follow it by voluntary choice: they "prefer" to carry out costly and often unpleasant actions to exclude, punish, and deter violators, and do not need to be threatened with exclusion, punishment or other deterrents to be induced to do so. More realistically, they may not actually "prefer" to act against violators, but wish to avoid disappointing the conventional expectations of fellow members of their group who rely on their help, and on whose help they wish to be able reciprocally to rely. (Splitting hairs, I am treating the threat of a sanction and the risk of disappointed reliance on an expected benefit, as different in kind. If they are not, the distinction I seek to make between spontaneous and enforced enforcement becomes blurred, and difficult to sustain.)

It is, however, stretching spontaneity beyond the breaking point to call an order spontaneous if it depends on "enforced enforcement," i.e., if members of the group or a subset of them punish and deter violators of the rules, not because they think it is in their reciprocal interest or simply because it is right to do so, but because they are threatened with exogenous sanctions if they fail to do it. In the latter case, with enforcement at one level depending on enforcement at the next higher level, who ultimately enforces enforcement?

Hayek is convinced that as civilization evolved, the scale of human coexistence changed by an order of magnitude, from small to great. There was a passage from the "face to face" society of small groups to the "Great Society" of the large group. Members of the small group were related to each other by ties of many kinds, and these relations gave rise to group solidarity. Members of the large group are unrelated and anonymous. They succeed to profit from the division of labor and the economies of scale made possible in their "extended order," not by relying on personal relations of trust, reciprocity and sympathy, but by respecting a suitable set of "abstract rules of just conduct."

Who, however, enforces the rules of just conduct? Respect for property and contract are not self-enforcing. On the contrary, they generate incentives for their own violation. Failing reliance on

reciprocity, there is no voluntary convention for enforcement, except for saying that in classical liberal doctrine it is the sole field where coercion is legitimate (1978, p. 109). In fact, if the Great Society works the way he believes it does, anonymously and at arm's length, enforcement cannot be supplied spontaneously, for it is undermined by the free rider problem and perverse incentives in exactly the same way as property, contract, and other prisoners' dilemmas. Enforcers must be coerced to enforce. Calling a spade a spade, one would say instead that enforcers must be paid to enforce, and for this to happen taxpayers must be coerced to pay taxes. There must be at the end of a regress of enforcement-enforcers, an ultimate, sovereign enforcer. There is no doubt whatever that when talking of the need to enforce rules of just conduct, it is the state that Hayek saw as the necessary, sufficient, and legitimate enforcer.

Here, too, vital pieces are missing from the jigsaw. Take the market order whose unique efficiency helps the large group to prevail. It is a web of exchanges of all kinds, most of them indefinitely repeated. In some instances, the two sides of the exchange are performed simultaneously. These are in most circumstances self-enforcing, and the parties to them might as well be anonymous (though usually they have names). When, however, performances are not simultaneous, executory contracts of some complexity are often required, and they are not self-enforcing. How could strangers with no name and no established reputation enter into such contracts with each other, the state's enforcing facilities notwithstanding? Do they ever do so? Who will be prepared to perform first when facing a nameless unknown? Yet, how, if you are anonymous, can you do any business at all except by performing first, unless somebody, broker, banker, insurance underwriter, middleman, lends you his name for love or money? In the Great Society, most people may well be anonymous to most others, because they have no profitable occasions to get acquainted; but since they have no such occasions, it does not matter much that they are anonymous. However, few people or none can remain anonymous to the handful of others with whom they interact in making the market order go round. That handful gets selected spontaneously, and it is always a "small group." There *is* no anonymous, large-group interaction because it would be too numerous to permit it. Its individual members interact in several "small groups" whose membership may be partly overlapping, partly different. Thus, each small group is open to other small groups and memberships are intermingled at the edges.

The truth of the matter about the Great Society is that few or no large groups are completely homogenous. Their membership can always be disaggregated, sorted into smaller groups by a variety of selection criteria. As the large group is always *the sum of small groups*, the converse goes as well: small groups can always be aggregated to form what is, from a chosen point of view, a large group. Its dimensions are in the eye of the beholder.

If one chooses to see it only as a large group, as Hayek does, something must be said about the existence problem of the non-spontaneous order which its specifications entail: if Hayek won't tell who enforced the enforcement, others will expound it with a vengeance.

For the effect of leaving out pieces from the jigsaw puzzle of social theory is that the vacuum is only too naturally filled by a false conception of the state. This conception is hardly compatible with liberal principles. Indeed, it is hardly compatible with the very market order that Hayek wants to be spontaneous, and culturally selected to make groups that adopt it succeed and groups that deliberately deform it, fail. For although it does not logically exclude other alternatives, Hayek's theory leads straight as an arrow to the facile conclusion of an indispensable state that alone upholds property and contract. They exist by the grace of society acting through the political authority. They function as society chooses that they should. The massive chorus we have been hearing from the left and center, chanting that property is a bundle of separable privileges granted or withheld by society, and the freedom of contract is subordinate to public policy, is vindicated by the very theory that should have prevailed over such a chorus with a clearer, a more powerful voice.

References

- Hayek, Friedrich A. von. 1960. *The Constitution of Liberty*. Chicago: University of Chicago Press.
- . 1973. *Law, Legislation, and Liberty*, Vol. 1, *Rules and Order*. Chicago: University of Chicago Press.
- . 1978. *New Studies in Philosophy, Politics, Economics and the History of Ideas*. London: Routledge and Kegan Paul.

Vertical Restraints and the Retail Free Riding Problem: An Austrian Perspective

David W. Boyd

Vertical restraints are restrictions that manufacturers place on the retailers who carry their products regarding to whom those dealers may resell the goods, or on the retail prices they may set. Although there exists a variety of economic explanations as to why rational manufacturers might wish to impose such downstream restrictions under different market conditions, perhaps the one which garners the most attention within the economics profession is that vertical restraints can help combat free riding problems within the retail sector.

The examination of vertical restraints as important pieces of the manufacturer's arsenal against free riding, however, like much of industrial organization based on neoclassical price theory, is steeped in equilibrium analysis. Ignored almost entirely in the literature is the dynamic nature of the free riding problem, and the critical roles that knowledge and discovery play in its evolution. Moreover, because the leading analyses consider a closed world of manufacturers and retailers in an isolated state of equilibrium, there exists no room for entrepreneurial activity.

In this article, I apply standard Austrian economic concepts—including the competitive market process and the roles played by knowledge, discovery, and the entrepreneur—to the analysis of vertical restraints in combating the retail free rider problem. The exposition is emphatically not meant as a thorough and exhaustive analysis of all vertical restraints from an Austrian viewpoint. Nor do I propose any new theoretical rationales for the employment of vertical restrictions. Rather, the paper

David W. Boyd is assistant professor of economics at Denison University.

The Review of Austrian Economics Vol. 9, No. 1 (1996): 119–34
ISSN 0889-3047

seeks to establish two points. First, an Austrian analysis of retail free riding provides meaningful insight into the problem beyond that incumbent in conventional neoclassical equilibrium models. Second, much of the often bitter debate over the welfare implications of vertical restrictions, by ignoring the market process, is substantially misguided. By presenting vertical restraints and retail free riding in an Austrian light, perhaps those who debate the welfare properties of vertical restrictions will begin incorporating the importance of the competitive market process into their arguments.

The first section of the paper briefly outlines the general retail free riding problem and discusses the standard analysis of how it can be fought with vertical restraints. The next section critiques the neoclassical analysis of vertical restrictions in combating free riding and applies the tools of Austrian economics to the problem. The third section addresses welfare, and the current policy debate over vertical restrictions. The final section provides commentary and a brief conclusion.

The Retail Free Riding Problem

A newly introduced, high quality product is not guaranteed market success. To compete successfully, the manufacturer must inform potential consumers of the product's existence, as well as supply at least a modicum of information regarding basic product characteristics, attributes, or capabilities. Although advertising can and often does effectively convey much of the most basic product information to potential consumers,¹ not all demand-enhancing, product-specific information can be efficiently provided by manufacturer advertising, or by other manufacturer actions. Rather, certain information is best imparted at retail establishments, including information gleaned from product demonstrations or especially knowledgeable salespersons. Because such pre-sale information is costly, the information-supplying retailer must recover the costs of information provision through the markup over its wholesale cost.

However, nothing prevents a retailer, or set of retailers, from opportunistically abandoning the pre-sale information services, thereby lowering cost, and luring consumers away with the resultant lower retail prices from the information-providing dealers.

¹See the thorough examination in Robert B. Ekelund and David S. Saurman, *Advertising and the Market Process: A Modern Economic View* (San Francisco: Pacific Research Institute for Public Policy, 1988).

That is, rational consumers can freely absorb the information offered by a "full-service" retailer, yet purchase from a "no-frills," discounting dealer to receive a lower retail price. In effect, these no-frills dealers can "free ride" on the information-supplying retailers. If this free riding activity continues unchecked, the full-service retailers, no longer able to recover their information costs due to reduced sales volume, will be forced to discontinue information service provision. Because these services are demand-enhancing, elimination of or reduction in the level of services offered at retail is harmful to the manufacturer.

Not all retailer-provided services are subject to this type of free riding. If the retailer is able to charge separately for the demand-enhancing services it offers, the additional costs associated with service provision can be adequately covered by a separate service price. For example, some specialty tennis retailers allow a customer to take one or more tennis racquet models home overnight to try out the equipment at the customer's leisure. If these high-service dealers charge separately for this service, as many do, then even if a customer discovers one model which she particularly fancies yet chooses to purchase it from a discounter at a reduced retail price, the specialty store will still be compensated for the service it offered. No free riding will have occurred.

Moreover, free riding will only be a problem to the extent that the retail services are offered *prior* to the sale. Post-sale services, since they are provided only to paying customers, are not subject to free riding. For instance, a home improvement store which provides substantial assistance to customers during the remodeling process *after the products have already been purchased* is likely to find free riding less of a problem than a retailer who only provides pre-sale services.

It is entirely possible that some subset of consumers will, over time, become fully informed and therefore find the information services offered by high-service retailers superfluous. Such customers will seek out discount retailers for their purchases. Nevertheless, so long as some not insignificant number of potential consumers still find retailer-provided information meaningful, free riding will remain a potential problem for those full-service dealers who provide complimentary pre-sale services.

The extent to which retail free riding is a pervasive and widespread problem is an open empirical question about which little has been written in the industrial organization literature. Free riding activity is not easily documented, and no systematic

inquiry has been made into the magnitude of the problem. The majority of the limited analysis which has been performed involves examination of individual litigated cases of vertical restraints.² Although the frequency and severity of retail free riding is uncertain, there is no question that the problem receives substantial attention in virtually every industrial organization textbook. If the resources of the profession devoted to the study of a particular problem serve as a proxy for its importance, then the retail free riding problem merits additional examination.

In any event, when retail free riding is of concern to a particular manufacturer, it has a clear incentive to undertake actions which can prevent, or at least ameliorate, the problem. One such set of actions is vertical restrictions.

Vertical Restraints and Retail Free Riding

It is the prospect of lower retail prices that entices consumers away from full-service retailers to discounting, no-frills dealers. The manufacturer can prevent such price cutting by imposing minimum resale price maintenance (RPM) on its dealer network.³ By setting a retail price floor below which no retailer may sell the

²See, for example, the discussion of the case involving Lenox china in V. Goldberg, "Enforcing Resale Price Maintenance: The FTC Investigation of Lenox," *American Business Law Journal* 18 (1980): 225-58; and H. P. Marvel and S. McCafferty, "Resale Price Maintenance and Quality Certification," *Rand Journal of Economics* 15 (1984): 346-59; or S. Oster, "Levi Strauss," in *Impact Evaluations of Federal Trade Commission Vertical Restraints Cases*, R. N. Lafferty, R. H. Lande, and J. B. Kirkwood, eds. (Washington, D.C.: Federal Trade Commission, 1984), an examination of the case of Levi Strauss jeans.

³Technically, RPM is a *per se* violation of the antitrust laws, under the standard set forth in *Dr. Miles Medical Company v. John D. Park & Sons Company*, 220 U.S. 373 (1911). The Supreme Court has, however, over the past few years, significantly narrowed the scope of *per se* illegal activity. See the Court's decisions in *Monsanto Company v. Spray-Rite Service Corporation* (1984) and *Business Electronics Corporation v. Sharp Electronics Corporation* (1988). In *Monsanto*, the Court ruled that a *per se* violation could not be inferred solely from evidence that a manufacturer had terminated a price-cutting distributor after having received complaints from the distributor's competitors. In *Sharp*, the Court went even further, holding that an agreement between a manufacturer and a dealer whereby the manufacturer would terminate any price-cutting dealer was not the same as an RPM agreement. Rather, *per se* illegality required an agreement between the manufacturer and all dealers. For a thorough legal history of RPM, see T. R. Overstreet, *Resale Price Maintenance: Economic Theories and Empirical Evidence* (Washington, D.C.: Federal Trade Commission, 1983); P. Ippolito, *Resale Price Maintenance: Economic Evidence from Litigation* (Washington, D.C.: Federal Trade Commission, 1988); or M. C. Katz, "Vertical Contractual Relations," in *Handbook of Industrial Organization*, Richard Schmalensee and Robert W. Willing, eds. (New York: North Holland, 1989).

product, the standard argument goes, the manufacturer deflects the ruinous price competition into non-price competition in information provision.

Of course, RPM can and is used for reasons other than combating the retail free riding problem. A number of other rationales for the practice have withstood scholarly scrutiny, and still receive substantive attention in the literature. For example, RPM may facilitate dealer and/or manufacturer collusion, serve as an aid to price discrimination, or provide an incentive device to promote the establishment of additional retail outlets.⁴ Nevertheless, the limited empirical record seems to suggest that, at least for litigated cases involving RPM, the practice is predominantly used to minimize free riding. Ippolito in her systematic examination of all public and private cases filed between 1976 and 1982 alleging vertical price restraints, finds that the collusion theories are plausible explanations for no more than 15 percent of all case filings, and an even smaller percentage of the private cases. Yet roughly 70 percent of the cases appear to be situations where retail free riding could have occurred. In fact, virtually all of the cases she examined are consistent with one or more of the "special services" theories of RPM.⁵

Resale price maintenance is not the only vertical restraint option open to the manufacturer. Alternatively, the manufacturer could remove any potential free riding activity by granting to a single dealer in each geographical market an exclusive territory. Since the retailer granted the exclusive territory is insulated from any potential free riders, the manufacturer's incentive to have information services provided is perfectly transferred to the dealer.

⁴A number of authors have written on these and other justifications for RPM. See, for example, the classic works of B. Yamey, "Origins of Resale Price Maintenance: A Study of Three Branches of Retail Trade," *Economic Journal* 62 (1952): 552 and *Resale Price Maintenance*, Yamey, ed. (Chicago: Aldine Publishing, 1966); or, more recently, G. F. Mathewson and R. A. Winter, "An Economic Theory of Vertical Restraints," *Rand Journal of Economics* 15 (1984): 27-38; or S. I. Ornstein, "Resale Price Maintenance and Cartels," *The Antitrust Bulletin* 35 (1985): 1-18; on the collusion theories, W. S. Bowman, "The Prerequisites and Effects of Resale Price Maintenance," *University of Chicago Law Review* 22 (1955): 825; or R. E. Caves, "Vertical Restraints as Integration by Contract: Evidence and Policy Implications," Harvard Institute of Economic Research Discussion Paper No. 754 (1980) on price discrimination; and J. R. Gould and L. E. Preston, "Resale Price Maintenance and Retail Outlets," *Economica* 302 (1965): 302-12 on the outlets rationale. For a thorough summary of all economic justifications for RPM, see Overstreet, *Resale Price Maintenance: Economic Theories and Empirical Evidence*; or Ippolito, *Resale Price Maintenance: Economic Evidence from Litigation*.

⁵Ippolito, *Resale Price Maintenance: Economic Evidence from Litigation*.

Neither of these vertical restraints, however, completely solves the manufacturer's free riding problem. Because consumers do not always make locational purchase decisions based solely on the quantity and quality of the information provided by dealers, the uniform retail prices guaranteed by RPM do not necessarily generate a pure incentive for dealers to provide the level of information manufacturers desire. In particular, residual free riding can occur, causing retailers to select a quantity of information provision strictly less than the manufacturer's preferred level.⁶ Since these information services are demand-enhancing, insufficient information investment by retailers reduces aggregate sales and hence manufacturer profitability.

On the other hand, the artificial elimination of retail competition spawned by the geographical restrictions associated with exclusive territories confers to the dealer sufficient market power to mark up retail prices accordingly. The higher retail prices constrict the manufacturer's derived demand, also reducing profitability.

The conventional analysis, then, suggests that the retail free riding problem must be fought with one of these two imperfect strategies. An Austrian insight into the problem and these two proposed remedies, however, reveals not only that the vertical restraint solutions are likely to be much less effective than the orthodox analysis allows, but that a third solution can, and often does, help mitigate the retail free riding problem.

Vertical Restraints and the Retail Free Riding Problem: An Austrian Critique

Knowledge

To compare the relative efficacy of RPM and exclusive territories in combating the retail free riding problem, and to implement the better strategy, requires that the manufacturer possess a set of market knowledge that is impressive in both its breadth and depth. Under RPM, the manufacturer must devise an optimal retail price, below which it desires no retailer to offer the product to consumers. To determine this retail price floor, the manufacturer must be thoroughly acquainted with the demand schedule for its product. But as Mises made clear,⁷ the manufacturer is

⁶See, David W. Boyd, "The Choice Between Resale Price Maintenance and Exclusive Territories," *Review of Industrial Organization* 8 (1993): 755-63, for a more extended analysis of residual free riding under RPM.

⁷Ludwig von Mises, *Human Action* (Chicago: Henry Regnery, 1966), p. 378.

unlikely to know the shape of its demand curve. This knowledge problem is multiplied dramatically under exclusive territories, where the manufacturer must know the demand for its product in *each separate geographical market* in order to efficiently design the territories themselves and set an appropriate wholesale price in each.

The knowledge problem is likely to be especially acute for the manufacturers of newly introduced products—precisely those products for which the information supplied by retailers is most critical. The assumption that a manufacturer of a new product never before marketed possesses sufficient knowledge of a *hypothetical* demand curve to make meaningful decisions about optimal wholesale and retail prices certainly strains the bounds of credibility.

Furthermore, for RPM to prevent sales from bleeding to lower priced outlets, the retail price floor set by the manufacturer must be uniform across the entire retail network. But a single retail price will be optimal only if all retailers incur identical costs, a popular assumption within the literature. However, the actual costs which affect retailer decision making are subjective opportunity costs,⁸ known only to each individual dealer. Even if the manufacturer were somehow able to magically learn of all these subjective retailer opportunity costs, it is exceedingly unlikely that they would all be exactly the same. In all actuality, no single optimal retail price exists for the manufacturer to set.

The Information Entrepreneur

The most often cited analyses of vertical restraints in fighting retail free riding consist of insulated manufacturers and retailers operating within a closed world. Solutions involving other parties are not considered as options. In an open universe of decision makers, however, the free riding problem for the manufacturer is, at the same time, an opportunity for an alert entrepreneur. Kirzner opens the entrepreneurial door when discussing the information content of advertising.

If producers did not offer “free” information jointly with the product, consumers would be forced to use other techniques of purchasing information. In fact, the provision of free information through advertising has, in this line of ideas, come more recently

⁸James M. Buchanan, “What Should Economists Do?,” *Southern Economic Journal* (1964): 213–223.

to be integrated into the broader theory of the economics of information. In this theory, the provision of information is treated as a service separated from the products to which market information is relevant. Moreover, it is a service whose usefulness is *valued* by the consumer *separately* from that of the relevant products.⁹ (emphasis in original)

Any reduction in information offered by full-service dealers as a response to free riding activity provides an opportunity for alert entrepreneurs to fill the information lacuna with their own information services. Since, as Kirzner pointed out above, product information is often valued by the consumer distinctly from the product itself, "information entrepreneurs" can bundle information packages and market them separately and directly to the consuming public. In fact, as argued below, the information supplied by third party information entrepreneurs is likely to be superior to that offered by retailers.

Consider, for example, the markets for bicycles and bicycle equipment. In the last decade or so, bicycles and bicycle equipment have become increasingly complex, due in part to specialization and advances in technology. Today, sporting goods dealers and specialty bicycle shops sell a broad assortment of racing bikes, road bikes, mountain bikes, and so called "hybrid" bikes, which combine features of racing and mountain bikes. Each type of bicycle requires its own assortment of specialized parts, components, and accessories. To make an intelligent purchasing decision, a consumer requires information about the various products available, as well as assistance in arriving at an appropriate type and model of bike, given the manner in which he or she is apt to use it. Much of this information is provided by full-service retailers via bicycle displays and specially trained salespeople. However, discount outlets and mail-order marketers often distribute many of the same products, without offering a commensurate level of information services, at lower retail prices. Consequently, free riding activity in this market is distinctly possible. Indeed, many manufacturers appear to attempt to combat the problem by, for example, setting "manufacturers' suggested retail prices" to discourage price cutting, and imposing a degree of territorial exclusivity in retail networks. Nevertheless, some free riding certainly does occur, with the result that the information

⁹Israel M. Kirzner, *Competition and Entrepreneurship* (Chicago: University of Chicago Press, 1973), p. 152.

services offered by full-service dealers are, in all likelihood, insufficient from both the manufacturer's and the consumer's point of view.

In this setting, a variety of information entrepreneurs have developed products and services to help fill the information gap spawned by retail free riding. For example, magazines, books, and cable television programming all provide consumers with information about bicycles and bicycling. A host of magazines appear on the shelves of any newsstand covering bicycle racing and mountain biking. For more general information, a variety of books describe the different kinds of bicycles available in the market today, and offer advice to consumers regarding which type of bicycle might be most appropriate. Sports channels on cable television regularly televise both road and mountain bike races. All of these media are potentially important sources of information for anyone imperfectly informed about the products offered in today's bicycle market.

The number and variety of information entrepreneurs will vary with the product's value and complexity. Consumers of expensive, complicated, and highly technological products are apt to demand more information than the purchasers of less inexpensive, mundane products. Though product information is important for potential buyers of both computers and running shoes, for example, the demand for computer information is likely to exceed in depth and breadth the demand for information about running shoes. Thus, while one can obtain information about running shoes and computers from different magazine publications, information entrepreneurs have developed an assortment of additional information services—e.g., seminars and other classroom offerings—directed toward computer purchasers.

The ability of third-party entrepreneurs to offer successful information services depends critically on how quickly the information they provide depreciates. If the manufactured products in question are relatively stable—they do not change significantly over time—then consumers will quickly reach the point where the marginal cost of additional information acquisition exceeds the marginal benefit, and any market for information services will soon evaporate. However, if the product of concern is in a steady state of flux—that is, if the product is regularly altered, or if new and improved versions are constantly appearing on the market—then information entrepreneurs can provide a regularly updated stream of information services which the information-consuming public will readily demand. Indeed, if one subscribes

to the market process view of imperfect competition,¹⁰ most manufacturers, seeking to better satisfy evolving consumer wants and desires, will constantly update their products' qualities and characteristics. This view suggests that, at least in many contexts, markets will persist for the services provided by information entrepreneurs.

Market Process

The most oft-cited and influential analyses of vertical restraints as ammunition against retail free riding are neoclassical equilibrium models. In addition to overlooking the roles played by knowledge and entrepreneurship, these analyses ignore the competitive market process so essential to a thorough understanding of the free riding problem. In reality, a consumer neither enters the market as the *tabula rasa* many models initially assume, nor remains in the fully informed state of final equilibrium many of these very same models eventually reach. Instead, a consumer's knowledge base is in a constant state of flux, as he discovers new information, and forgets some as well. Neither is the relevant pool of information stable. New knowledge is constantly available, and other information becomes obsolete. To truly understand the retail free riding problem, one must at least acknowledge the dynamic process of knowledge acquisition.

Manufacturers themselves often respond to the evolving state of consumer information. Recognizing that many of their traditional customers have, over time, accumulated a body of knowledge about their products and services sufficient to make much of the information provided by traditional retailers unnecessary, Apple Computer, Inc. recently expanded their operations into the mail-order business, distributing a catalog offering for sale a wide variety of Apple hardware and software products. The company also seemed to address those consumers at the opposite end of the knowledge-base spectrum—people with little aversion to remaining relatively ignorant of personal computers and their workings. Apple's Performa™ line of machines, bundled with special software which make the computers easier, but less flexible, to use are distributed through discount outlets and department stores. Both of these distribution expansions reflect changes in the knowledge bases of consumers, and in the process by which

¹⁰Ibid., pp. 112–19.

they acquire information. Neither move enamored Apple with its traditional retailers, but market forces required the company to respond to the changing states of consumers' knowledge and preferences.

Information entrepreneurs must compete against one another, as well as with the information services provided by retailers and/or manufacturers. In the competitive market process, entrepreneurs offer a constantly changing information package, trying to better serve consumer demands and responding to evolving consumer knowledge bases. In many ways, the information provided by third party information entrepreneurs is likely to be superior to that offered by traditional retailers. While the information supplied by retailers is primarily ancillary to the actual product being distributed, the product being marketed by the information entrepreneur is the *information itself*. A retailer might consummate a transaction even if the information services offered there are quite poor. Information entrepreneurs are not so fortunate. Because the product *is* the information, inferior quality information translates directly into reduced sales. Furthermore, because the services offered by entrepreneurs often include information on the products of a number of manufacturers, the information is less likely to be biased than that offered by retailers, who many have an incentive to steer consumers towards those brands for which the dealer receives the largest retail margin.

Moreover, the competition between traditional retailers usually is constrained to some extent by transportation costs. Dealers generally only compete against other retailers within a relatively small geographic proximity. In the extreme case of exclusive territories, competition between dealers is virtually eliminated. Information entrepreneurs, on the other hand, are typically able to compete in a larger range of markets, since their products are often broadly available via bookstores, newsstands, mail services, electronic networks, or public broadcast spectrums. Because the competition between information entrepreneurs is more intense than the competition between retailers, the information products supplied by third-party entrepreneurs are likely to be of a higher quality.

This is not to say that all information is best provided by information entrepreneurs. Rather, the competitive market process will tend to impose efficiency in information distribution. At any point in time, the types of information most productively provided by retailers will tend to be supplied by them. Market forces

will discipline entrepreneurs to offer the information services that they can more efficiently supply. And the mix between all types of information providers will constantly sway, reflecting changes in consumer demands and the state of consumer knowledge. In short, the competitive market process will determine which information will be provided by whom.

The Welfare Debate

Even though both RPM and exclusive territories can at least partially solve the retail free riding problem, they are currently subject to different legal treatments. Vertical price restrictions, including RPM, are technically *per se* violations of the antitrust laws, although manufacturers may legally engage in activity which is nearly indistinguishable from RPM.¹¹ Vertical non-price restraints, including exclusive territories, are presently judged under the more lenient rule-of-reason standard.¹² Much of the public policy debate over the retail free riding problem has concentrated on this disparity in legal treatments, and, in particular, on the seemingly overly stringent *per se* standard applied to RPM.¹³

Central in this debate is the effect RPM has on retail prices and, therefore, social welfare. Opponents of RPM argue that the practice, by artificially precluding price reductions, raises retail prices to consumers and therefore reduces welfare. RPM proponents counter that the larger retail margin, necessary for retailers to pay for information services, comes from lower wholesale prices, not higher retail prices. Furthermore, because RPM increases the information available to consumers via retailers, welfare is enhanced.

Using prices as a barometer for welfare and efficiency, however, requires an interpersonal comparison between the benefits and costs to consumer and producer. Since these benefits and costs are subjective, such a comparison is meaningless. Buchanan provides a more sensible approach to analyzing the welfare properties of vertical restraints and other activities: any activity,

¹¹See note 3 above.

¹²The current legal standard to be applied to vertical non-price restraints is set forth in *Continental TV, Inc. v. GTE Sylvania, Inc.*, 433 U.S. 36 (1977).

¹³See the exchange between David W. Boyd, "The Resale Price Maintenance Struggle: A Comment," *American Journal of Economics and Sociology* (October 1993): 447-54 and Anthony J. Greco, "Response of Professor Anthony J. Greco," *American Journal of Economics and Sociology* (October 1993): 454-58.

institution, or legal rule which facilitates mutually advantageous, voluntary exchange is efficient, or welfare-enhancing.¹⁴

The information entrepreneur measures up particularly well using Buchanan's welfare yardstick. When the retail free riding problem is solved by vertical restraints, and information is only provided by dealers, the consumer is forced to pay for information services, via the elevated retail price, whether he or she desires them or not. By unbundling the information from the product, the information entrepreneur makes the information available to the consumer without forcing all consumers to pay for information dissemination. Those consumers who demand information services are free to purchase them. Those who prefer to remain rationally ignorant or who already possess sufficient knowledge are not forced to subsidize information collection and dissemination.

Furthermore, by marketing the information products directly to the consumer, the information entrepreneur is likely to reduce consumer search costs, and therefore the effective price paid by consumers. Information provided by retailers nearly always necessarily entails a trip to the dealer. The transportation costs and the opportunity costs of the time forgone to acquire this information can be substantial, especially when a number of dealers are visited. Information entrepreneurs, on the other hand, typically deliver the information services directly to the consumer. Although they entail their own prices, acquisition costs are apt to be minor.

The information entrepreneur, therefore, is clearly a welfare-enhancing actor. Yet his place in the welfare debate over the retail free riding problem is notoriously absent. The welfare discussion should be expanded to include the roles played by the information entrepreneur and the institutions that affect how well he can perform his tasks. For example, the establishment of property rights in information influences the types of products and services the information entrepreneur can provide in a competitive market. By more concretely defining such rights, the third-party information provider can develop his products in a less uncertain environment. In general, attention in the debate should be paid to institutions and legal rules which affect the ability of the information entrepreneur to engage in his welfare-enhancing activities.

¹⁴James M. Buchanan, "Introduction: L. S. E. Cost Theory in Retrospect," in *L. S. E. Essays on Cost*, Buchanan and G. F. Thirlby, eds. (New York: New York University Press, 1981).

Conclusion

Despite a revival of Austrian economics since the middle 1970s, the school has made insufficient progress in injecting its ideas into mainstream economic thought. In a recent paper expanding on his work on the competitive market process, Kirzner conveys the sentiments of many Austrians: "Although this fundamental difference (between the perfectly competitive model and the competitive market process) has been articulated by the Austrians now for several decades, there has been disappointingly little impact upon mainstream contemporary theory."¹⁵ This failure may be due in part to an overly zealous concentration on theory at the expense of application. Indeed, W. Duncan Reekie largely agrees with this lament in his commentary on Kirzner's article,

students want either to study industrial organization and normative implications thereof for public policy toward business, or they want to study managerial economics to obtain normative insights on how to maximize profits. . . . To a degree, this coupling of student demands into "relevancy" and "low cost" may be mutually inconsistent; but that is the nature of the market and if we fail to meet this demand we, of all people, should not be surprised if others step in to fill it. . . . I believe Austrians have been frightened that . . . they will fling the baby out with the bathwater, but, in fear of doing so, the bath has gone cold and the baby is freezing to death.¹⁶

In this paper, I have attempted to respond to this criticism by applying well established Austrian economic notions to retail free riding, a significant problem in industrial organization. Examining the problem from an Austrian perspective reveals that the traditional neoclassical analysis ignores many of the most salient characteristics of the issue, and that much of the public policy debate on retail free riding is substantially misguided. If Austrian economists continue to illustrate that their

¹⁵Israel M. Kirzner, "The Driving Force of the Market: The Idea of 'Competition' in Contemporary Economic Theory and in the Austrian Theory of Market Process," in *Austrian Economics: Perspectives on the Past and Prospects for the Future*, Richard M. Ebeling, ed. (Hillsdale, Mich.: Hillsdale College Press, 1991), p. 139.

¹⁶W. Duncan Reekie, "The Driving Force of the Market: The Idea of 'Competition' in Contemporary Economic Theory and in the Austrian Theory of Market Process: Commentary," in *Austrian Economics: Perspectives on the Past and Prospects for the Future*, Richard M. Ebeling, ed. (Hillsdale, Mich.: Hillsdale College Press, 1991), pp. 169–70.

principles provide insight into everyday economic applications beyond that generated by mainstream economic thinking, perhaps the Austrian school will eventually make much needed inroads into mainstream economic thought.

References

- Bowman, W. S., Jr. 1955. "The Prerequisites and Effects of Resale Price Maintenance." *University of Chicago Law Review* 22: 825.
- Boyd, David W. 1993a. "The Choice Between Resale Price Maintenance and Exclusive Territories." *Review of Industrial Organization* (November): 755–63.
- . 1993b. "The Resale Price Maintenance Struggle: A Comment." *American Journal of Economics and Sociology* (October): 447–54.
- Buchanan, James M. 1981. "Introduction: L. S. E. Cost Theory in Retrospect." In *L. S. E. Essays on Costs*. James M. Buchanan and G. F. Thirlby, eds. New York: New York University Press.
- . 1964. "What Should Economists Do?," *Southern Economic Journal* (January): 213–22.
- Caves, R. E. 1980. "Vertical Restraints as Integration by Contract: Evidence and Policy Implications." Harvard Institute of Economic Research Discussion Paper No. 754.
- Ekelund, Robert B. and David S. Saurman. 1988. *Advertising and the Market Process: A Modern Economic View*. San Francisco: Pacific Research Institute for Public Policy.
- Goldberg, V. 1980. "Enforcing Resale Price Maintenance: The FTC Investigation of Lenox." *American Business Law Journal* 18: 225–58.
- Gould, J. R. and L. E. Preston. 1965. "Resale Price Maintenance and Retail Outlets." *Economica* 302: 302–12.
- Greco, Anthony J. 1993. "Response of Professor Anthony J. Greco." *American Journal of Economics and Sociology*. October: 454–58.
- Ippolito, P. 1988. *Resale Price Maintenance: Economic Evidence from Litigation*. Washington, D.C.: Federal Trade Commission.
- Katz, M. C. 1989. "Vertical Contractual Relations." In *Handbook of Industrial Organization*. Richard Schmalensee and Robert W. Willig, eds. New York: North Holland.
- Kirzner, Israel M. 1992. *The Meaning of Market Process: Essays in the Development of Modern Austrian Economics*. London: Routledge.
- . 1991. "The Driving Force of the Market: The Idea of 'Competition' in Contemporary Economic Theory and in the Austrian Theory of Market Process." In *Austrian Economics: Perspectives on the Past and Prospects for the Future*. Hillsdale, Mich.: Hillsdale College Press.
- . 1973. *Competition and Entrepreneurship*. Chicago: University of Chicago Press.

- Marvel, H. P. and S. McCafferty. 1984. "Resale Price Maintenance and Quality Certification." *Rand Journal of Economics* 15: 346–59.
- . 1986. "The Political Economy of Resale Price Maintenance." *Journal of Political Economy* 94: 1074–95.
- Mathewson, G. F. and R. A. 1984. "An Economic Theory of Vertical Restraints." *Rand Journal of Economics* 15: 27–38.
- Mises, Ludwig von. 1966. *Human Action*. Chicago: Henry Regnery.
- Ornstein, S. E. 1985. "Resale Price Maintenance and Cartels." *The Antitrust Bulletin* 35: 1–18.
- Oster, S. 1984. "Levi Strauss." In *Impact Evaluations of Federal Trade Commission Vertical Restraints Cases*. R. N. Lafferty, R. H. Lande, and J. B. Kirkwood, eds. Washington, D.C.: Federal Trade Commission.
- Overstreet, T. R. 1983. *Resale Price Maintenance: Economic Theories and Empirical Evidence*. Washington, D.C.: Federal Trade Commission.
- Reekie, W. Duncan. 1991. "The Driving Force of the Market: The Idea of 'Competition' in Contemporary Economic Theory and in the Austrian Theory of Market Process: Commentary." In *Austrian Economics: Perspectives on the Past and Prospects for the Future*. Richard M. Ebeling, ed. Hillsdale, Mich.: Hillsdale College Press.
- Rey, P. and J. Tirole. 1986. "The Logic of Vertical Restraints." *American Economic Review* 76: 921–39.
- Telser, L. 1960. "Why Should Manufacturers Want Fair Trade?" *Journal of Law and Economics* 3: 86–105.
- Thomsen, E. F. 1992. *Prices and Knowledge: A Market Process Perspective*. London: Routledge.
- Yamey, B. 1952. "Origins of Resale Price Maintenance: A Study of Three Branches of Retail Trade." *Economic Journal* 62: 552.
- , ed. 1966. *Resale Price Maintenance*. Chicago: Aldine Publishing.

Notes and Replies

Rejoinder: Salerno on Calculation, Knowledge, and Appraisalment

Leland B. Yeager

Contrary to Joseph Salerno's "Reply" (1994), my article (1994) did not try to specify what Professors Salerno, Rothbard, and Herbener mean by "calculation." I did not say that they "equate calculation and arithmetic" (Salerno 1994, p. 116). Instead, I implicitly raised the *question* of what they might mean. I did so by pursuing my avowed main purpose, namely, to challenge their attempted dichotomy between the "calculation" and "knowledge" problems of socialism. I further challenged their suggestion that Ludwig von Mises was concerned only with the non-knowledge aspects of the calculation problem.

Salerno now focuses attention on "appraisalment." In the capitalist market process but not under socialism, every scarce resource receives an appraisalment "reflecting its relative importance in serving (anticipated) consumer preferences" (p. 112). The market process "transforms the substantially qualitative knowledge about economic conditions acquired . . . by competing entrepreneurs," including their estimates of consumers' subjective valuations, into "an integrated system of objective exchange ratios for the myriads of original and intermediate factors of production" (ibid.). The process operates "in the face of ceaseless change of the economic data" (p. 113). "[D]ispersed knowledge is not a bane but a boon to the human race; without it, there would be no scope for the intellectual division of labor, and social cooperation under division of labor would, consequently, prove impossible" (pp. 114–15). In the capitalist process, as Rothbard says in a passage that Salerno quotes (p. 115), "the entrepreneur is able to appraise and estimate future costs and prices. In the

Leland B. Yeager is professor emeritus of economics at Auburn University.

The Review of Austrian Economics Vol. 9, No. 1 (1996): 137–39
ISSN 0889-3047

market economy, *qualitative* knowledge can be transmuted, by the free price system, into rational economic calculation of *quantitative* prices and costs."

Correctly interpreting Mises, Salerno writes "that the market creates a social appraisalment process . . . which depends crucially on an intellectual division of labor featuring the qualitative *understanding* of competing entrepreneurs; and . . . that this process is indispensable for converting the multidimensional knowledge of the economic data, regardless of who possesses this knowledge or where it is located, into a unitary structure of meaningful resource and product prices" (p. 120). Salerno goes on to emphasize "entrepreneurial forecasting" and "dynamic appraisalment" (p. 122).

All this emphasis on knowledge—yes, knowledge—is amply warranted. A further point that Salerno makes at least implicitly is worth underlining: much relevant knowledge is of a kind that simply could not be quantified or articulated except through the market process. Often an entrepreneur makes business decisions partly on his intuition or feel for technology, the attitudes and tastes of consumers and workers, sources of financing, and conditions in markets for inputs and for consumer goods and services—all in the future as well as the present. The entrepreneur receives information for judgments about such matters by reading specialized and popular publications, watching television and movies, experiencing various services and products personally, chatting with innumerable people, and strolling through town or the shopping mall. Much of what he thereby observes—or senses—he could not express in explicit words or numbers. A socialist system would let much such entrepreneurial knowledge go to waste even if it somehow emerged in the first place. The market system of profit and loss tends, as Mises said, to expand or shrink the roles of particular entrepreneurs according as they are relatively good or relatively poor in acquiring and using even these subtle and conjectural kinds of knowledge.

I forgo correcting several of Salerno's conjectures about my own position, mind-set, and standards of scholarship because his conjectures are peripheral to the issue at hand. My central point is that Mises was indeed concerned with the capitalist market process of mobilizing, articulating, and quantifying knowledge that socialism simply could not duplicate or replace. On any reasonable interpretation of exactly what calculation means in debates over socialist calculation, it is closely bound up with the development and use of knowledge. One ill-serves

Mises's reputation and ill-serves wide understanding of momentous issues by insisting on some conception or other of calculation from which all knowledge aspects have been severed.

References

- Salerno, Joseph T. 1994. "Reply to Leland B. Yeager on 'Mises and Hayek on Calculation and Knowledge'." *Review of Austrian Economics* 7, no. 2: 111–25.
- Yeager, Leland B. 1994. "Mises and Hayek on Calculation and Knowledge." *Review of Austrian Economics* 7, no. 2: 93–109.

A Final Word: Calculation, Knowledge, and Appraisalment

Joseph T. Salerno

Leland Yeager's rejoinder "Salerno on Calculation, Knowledge, and Appraisalment" (1996) merits a final word, because I believe it makes a subtle though very important concession to the argument made by Rothbard, Herbener, and myself regarding the function of the price system. Before addressing this issue, I will offer comments relating to two other points raised by Yeager in his rejoinder.

1. Yeager (p. 175) denies having said, in his original article (1994), that Salerno, Rothbard, and Herbener (SRH) "equate arithmetic and calculation." While strictly speaking he is correct, I believe it is a quite reasonable inference from the statements he does make when characterizing the problem confronting central planners who are endowed with perfect knowledge of the economic data. As I have discussed in my Reply (Salerno 1994, pp. 120–23), in SRH's interpretation of Mises's argument, granting the unrealistic assumption of complete and perfect information, about either present disequilibrium market conditions or the future final equilibrium state toward which the real market is tending at any moment, does not obviate the need for monetary calculation provided by the market's pricing process. Yeager (1994, p. 97) states, however, that given "the centralized availability of information," the argument against socialism "boil(s) down to a contention about arithmetic." Elsewhere, in questioning the "supposed distinction" that SRH observe in Mises's writings between calculation and knowledge, Yeager (1994, p. 94) remarks "I cannot believe Mises was merely saying that *if* the socialist planners possessed in some way all the information normally conveyed by the market prices, they still would be stymied

Joseph T. Salerno is associate professor of economics at the Lubin School of Business, Pace University.

The Review of Austrian Economics Vol. 9, No. 1 (1996): 141–42
ISSN 0889-3047

by inability to perform *calculations* in the narrow arithmetic sense." What else could Yeager mean by these statements except that anyone who argues that Mises distinguished calculation from knowledge is indeed equating calculation with arithmetic?

2. In stating that "Salerno *now* focuses attention on 'appraisement' (the emphasis is mine)" Yeager is implying that this is a concept I originated in my Reply (Salerno 1994) and that I have, therefore, suddenly shifted the grounds of my initial argument. But in my Reply (*ibid.*, pp. 112–15) I quote at length several passages from my earlier work on this topic, as well as from Rothbard's, in which the concept of "social appraisement" is elaborated. Far from an *ad hoc* fallback position, the notion of appraisement was at the heart of the SRH interpretation of Mises from the very beginning.

3. Finally, and most importantly, in the two consecutive paragraphs ending with the penultimate paragraph of his rejoinder, Yeager now abandons his original Hayekian position that the price system, i.e., past prices, automatically conveys to all passive producers all the knowledge that is relevant to their business decisions in a near-equilibrium world. He now concedes that "knowledge" is a primary matter of individual entrepreneurial experience, hunches, reading and personal observation, and superior forecasting of the uncertain future and whose qualitative content is therefore *not* "normally conveyed by (past) market prices." Contrary to his claim in the final paragraph of his rejoinder, then, this means that Mises's concept of economic calculation refers exclusively to the function of the price system in permitting entrepreneurs to appraise the quantitative importance of productive resources in a world where incessant change renders the future very unlike the past and current prices not directly relevant to future-oriented production plans.

References

- Salerno, Joseph T. 1994. "Reply to Leland B. Yeager on 'Mises and Hayek on Calculation and Knowledge.'" *Review of Austrian Economics* 7, no. 2: 111–25.
- Yeager, Leland B. 1994. "Mises and Hayek on Calculation and Knowledge." *Review of Austrian Economics* 7, no. 2: 93–109.
- . 1995. "Salerno on Calculation, Knowledge, and Appraisement." *Review of Austrian Economics* 9, no. 1: 175–77.

Socialism: A Property or Knowledge Problem?

Hans-Hermann Hoppe

In a series of recent articles in *The Review of Austrian Economics*, Joseph Salerno began to de-homogenize the often conflated economic and social theories of Ludwig von Mises and Friedrich A. Hayek. In particular, he has shown that their views on socialism are distinctly different, and he has argued in effect that Mises's original argument in the so-called socialist calculation debate was correct all along and was also the final word, whereas Hayek's distinct contribution to the debate was fallacious from the outset, and merely added confusion. The following note will provide additional support to Salerno's thesis.

Mises's well-known calculation argument states this: If there is no private property in land and other production factors, then there can also be no market prices for them. Hence, economic calculation, i.e., the comparison, in light of current prices, of anticipated revenue, and expected cost expressed in terms of a common medium of exchange—money—(thus permitting *cardinal* accounting operations), is literally impossible. Therefore, socialism's fatal error is the absence of private property in land and production factors, and, by implication, the absence of economic calculation.

For Hayek, socialism's problem is not a lack of property but a lack of knowledge. His distinctly own thesis is altogether different from Mises's.¹ For Hayek, the ultimate flaw of socialism is the fact that knowledge, in particular "the knowledge of the

Hans-Hermann Hoppe is professor of economics at the University of Nevada, Las Vegas.

¹See in particular the widely acclaimed 1945 article on "The Use of Knowledge in Society," reprinted in F. A. Hayek, *Individualism and Economic Order* (Chicago: University of Chicago Press, 1948).

The Review of Austrian Economics Vol. 9, No. 1 (1996): 143–49
ISSN 0889-3047

particular circumstances of time and place," exists only in a widely dispersed form as the personal possession of various individuals; hence, it is *practically* impossible to assemble and process all the actually existing knowledge within the mind of a single socialist central planner. Hayek's solution is not private property, but the decentralization of the use of knowledge.

Yet this is surely an absurd thesis. First, if the centralized use of knowledge is the problem, then it is difficult to explain why there are families, clubs, and firms, or why they do not face the very same problems as socialism. Families and firms also involve central planning. The family head and the owner of the firm also make plans which bind the use other people can make of their private knowledge, yet families and firms are not known to share the problems of socialism. For Mises, this observation poses no difficulty: under socialism private property is *absent*, whereas individual families and private firms are *based* on the very *institution* of private property. But for Hayek the smooth operation of families and firms is puzzling, because his idea of a fully decentralized society is one in which each person makes his own decisions based on his own unique knowledge of the circumstances, unconstrained by any central plan or supraindividual (social) norm (such as the institution of private property).

Second, if the *desideratum* is merely the decentralized use of knowledge in society, then it is difficult to explain why the problems of socialism are fundamentally different from those encountered by any other form of social organization. *Every* human organization, composed as it is of distinct individuals, constantly and unavoidably makes use of decentralized knowledge. In socialism, decentralized knowledge is utilized no less than in private firms or households. As in a firm, a central plan exists under socialism; and within the constraints of this plan, the socialist workers and the firm's employees utilize their own decentralized knowledge of circumstances of time and place to implement and execute the plan. For Mises, all of this is completely beside the point. But within Hayek's analytical framework, no difference between socialism and a private corporation exists. Hence, there can also be no more wrong with the former than with the latter.

Clearly, Hayek's thesis regarding the central problem of socialism is nonsensical. What categorically distinguishes socialism from firms and families is *not* the existence of centralized knowledge or the lack of the use of decentralized knowledge, but rather the absence of private property, and hence, of prices. In

fact, in occasional references to Mises and his original calculation argument, Hayek at times appears to realize this, too. But his attempt to integrate his very own thesis with Mises's and thereby provide a new and higher theoretical synthesis fails.

The Hayekian synthesis consists of the following propositional conjunction: "Fundamentally, in a system in which the knowledge of the relevant facts is dispersed among many people, prices can act to coordinate the separate actions of different people" and "the price system" can serve as "a mechanism for communicating information."² While the second part of this proposition strikes one as vaguely Misesian, it is anything but clear how it is logically related to the first, except through Hayek's elusive association of "prices" with "information" and "knowledge." However, this association is more of a semantic trick than rigorous argumentation. On the one hand, it is harmless to speak of prices as conveying information. They inform about past exchange ratios. But it is a non-sequitur to conclude that socialism's central problem is a lack of *knowledge*. This would only follow if prices actually *were* information. However, this is not the case. Prices convey knowledge, but they *are* the exchange ratios of various goods, which result from the voluntary interactions of distinct individuals *based on the institution of private property*. Without the institution of private property, the information conveyed by prices simply does not exist. Private property is the necessary condition—*die Bedingung der Möglichkeit*—of the knowledge communicated through prices. But then it is correct only to conclude, as Mises does, that it is the absence of the institution of private property which constitutes socialism's problem. To claim that the problem is a lack of knowledge, as Hayek does, is to confuse cause and effect, or premise and consequence.

On the other hand, Hayek's identification of "prices" and "knowledge" involves a deceptive equivocation. Not only does Hayek fail to distinguish between what one might call institutional knowledge—information that requires for its existence an institution (such as the knowledge of prices requires private property)—and *raw* or *extra-institutional knowledge*—like this is an oak tree, I like peanuts, or birds can fly. Moreover, Hayek also fails to notice that the knowledge of prices is not at all the same sort of knowledge whose existence he believes to be responsible for the "practical impossibility" of socialism and central

²Ibid., pp. 85–86.

planning. What makes central planning impossible, according to Hayek, is the fact that part of human knowledge exists only as essentially *private* information:

practically every individual has some advantage over all others because he possesses unique information of which beneficial use might be made, but of which use can be made only if the decisions depending on it are left to him or are made with his active cooperation.³

While it is certainly true that such knowledge exists, and while it is also true that uniquely private knowledge can of course never be centralized (without information losses), it is just as certainly *not* true that the knowledge of *prices* falls into this category of uniquely *private* information. To be sure, prices are “prices paid at specific times and places,” but this does not make them private information in the Hayekian sense. To the contrary, the information conveyed by *prices* is *public* information, because prices—*qua* objective exchange ratios—are *real* events. It may be *difficult* to know all of the prices paid at a specified date and location, just as it may be difficult to know every person’s physical location at any given time. Yet it is hardly *impossible* to know either one, and with current computer technology it is probably easy. In any case, while I may never know everything that you know, and *vice versa*, it is no more problematic to assume that both of us can simultaneously possess the same *price* information than that we can both simultaneously know the same baseball results. Hence, the knowledge conveyed by *prices* actually *can be centralized*. But if price information is public information and thus can be centralized, then, according to Hayek’s thesis that socialism’s problem stems from the inefficiency of trying to centralize genuinely uncentralizable private knowledge, it would follow that the absence of prices, and hence of private property, has *nothing to do* with the plight of socialism. Otherwise, if one insists with Mises that the absence of private property and prices *does* have something to do with the plight of socialism. Hayek’s contribution to the socialism debate must be thrown out as false, confusing, and irrelevant.

Hayek’s misconception of the nature of socialism is symptomatic of a fundamental flaw in his thinking, pervading not only

³Ibid., p. 80.

his economics but in particular also his political philosophy: his ultra-subjectivism. Hayek, as noted and quoted *ad nauseam* by his numerous followers, was convinced that “it is probably no exaggeration to say that every important advance in economic theory during the last hundred years was a further step in the consistent application of subjectivism.”⁴ While this may well be true, it does not logically follow that every further advance toward subjectivism must also lead to an advance in economic theory. However, Hayek seems to have drawn this conclusion and has thus become a prime example illustrating its falsehood.

Mises, and in his steps even more clearly Murray N. Rothbard, conceives of economics as the science of human *action*. Action has two inseparable aspects: a subjective aspect (action is rational, intelligible action) and an objective aspect (acting is always acting with real things and physical stuff). Accordingly, Mises’s and Rothbard’s economics and political philosophy is never anything but robust, and their categories and theories invariably possess real, operational meaning: private property, division of labor based on private property, production, direct and indirect exchange, and also compulsory interference with private property and production and exchange such as taxation, counterfeiting, legislation, and regulation.

In distinct contrast, Hayek—and misled by him to different degrees also Israel Kirzner and Ludwig Lachmann—views economics as some sort of science of human *knowledge*. Accordingly, Hayek’s categories and theories refer to purely subjective phenomena and are invariably elusive or even illusory. He is not concerned about acting with things, but about knowledge and ignorance, the division, dispersion, and diffusion of knowledge, alertness, discovery, learning, and the coordination and divergence of plans and expectations. The external—physical—world and real—material—events have almost completely disappeared from his view. Hayek’s categories refer to *mental* states of affairs and relationships, completely detached from and compatible with any real physical state of affairs and events.

Most notable and disturbing is the ultra-subjectivist turn in Hayek’s political philosophy. According to a long-standing tradition of political philosophy shared by Mises and Rothbard, freedom is defined as the freedom to privately own—and control—real property, and coercion is the initiation of physical

⁴*The Counterrevolution of Science* (New York: Free Press, 1955), p. 31.

damage—harm—upon the private property of others. In distinct contrast, Hayek defines freedom as “a state in which each can use his own *knowledge* and for his own *purposes*,”⁵ and coercion means “such control of the environment or circumstances of a person by another that, in order to avoid greater evil, he is forced to act not according to a coherent *plan* of his own but to serve the *ends* of another,”⁶ or alternatively, “coercion occurs when one man’s actions are made to serve another man’s will, not for his own but for the other’s *purpose*”⁷ (all emphases are mine). Clearly, Hayek’s definition contains nothing regarding scarce goods and real tangible property, and he provides no *physical* criterion or indicator whatsoever for the existence or non-existence of either state of affairs. Rather, coercion and freedom refer to specific configurations of subjective wills, plans, thoughts, or expectations. As *mental* predicates, Hayek’s definitions of freedom and coercion are compatible with every *real*, physical state of affairs. They possess no power to make any real distinctions.⁸

It is beyond the scope of this note to offer a detailed critique and refutation of Hayek’s ultra-subjectivism. However, beside the fundamental question whether a science of knowledge as envisioned by Hayek is even possible, i.e., whether there can be any other science of knowledge apart from logic and epistemology on the one hand and the history of ideas on the other,⁹ two conclusions are painfully clear. Even if Hayek’s science of knowledge is possible, it appears at best irrelevant because it is praxeologically—operationally—meaningless. At worst it is intellectually pernicious in promoting relativism.

As for the real world of acting with physical property, of production and exchange, of money and markets, profits and losses, capital accumulation and bankruptcies, there can be no lasting doubt about the existence of *laws* and the ceaseless operation of a *tendency* toward general equilibrium—action-coordination. Likewise, there can be no doubt about the existence of laws and the constant operation of dis-equilibrating tendencies within

⁵*Law, Legislation, and Liberty*, Vol. 1 (Chicago: University of Chicago Press, 1973), pp. 55–56.

⁶*Constitution of Liberty* (Chicago: University of Chicago Press, 1960), pp. 20–21.

⁷*Ibid.*, p. 133.

⁸See also Hans-Hermann Hoppe, “Hayek on Government and Social Evolution,” *Review of Austrian Economics* 7, no. 1 (1994): esp. 70f.

⁹For some serious doubts on this see Hans-Hermann Hoppe, *Kritik der kausalwissenschaftlichen Sozialforschung* (Opladen: Westdeutscher Verlag, 1983).

the world of actual taxation, counterfeiting, legislation, and regulation. Indeed, it would be extremely costly—prohibitive—to not recognize such laws and tendencies and to adopt relativistic views. In contrast, in surreptitiously shifting attention from the tangible world of action and property to the ethereal world of knowledge, ideas, plans and expectations, relativistic views become attractive (and cheap). There are no apparent regularities and tendencies in Hayek's knowledge world. In fact, it is difficult to even imagine what "law" and "equilibrium" could possibly mean in the context of purely subjective phenomena. Instead there exists seemingly nothing but constant kaleidoscopic change.

It is hardly surprising, then, that Hayek and his followers could proclaim such relativistic slogans as that we cannot do anything to improve our condition except rely on spontaneous evolution, that our future is completely unknowable, or that we cannot but participate in an endless and open-ended stream of conversation. As far as the realm of purely subjective phenomena is concerned, and as addressed to a purely spiritual—disembodied—being, this may well be good advice. Actually possessing physical—bodily—existence, however, why would anyone even care to know it? As applied to the world of bodily action and property, such advice is self-destructive nonsense.

Calculation and the Question of Arithmetic

Jeffrey M. Herbener

The view that Ludwig von Mises had more in mind in his calculation critique of socialism than the Hayekian knowledge problem has recently been attacked by Leland Yeager.¹ This article addresses Yeager's central claim that,

I cannot believe Mises was merely saying that *if* the socialist planners possessed in some remarkable way all the information normally conveyed by genuine market prices, they still would be stymied by inability to perform *calculations* in the narrow arithmetical sense, an inability that advances in supercomputers might conceivably overcome.²

Yeager then asserts that Joseph Salerno, Murray Rothbard, and I (SRH) claim that this is what Mises meant. If Yeager means by this assertion that we believe that this is Mises's entire calculation argument, then Salerno is correct in responding that,

Jeffrey M. Herbener is associate professor of economics at Washington and Jefferson College and wishes to thank the anonymous referees for their helpful comments.

¹Leland Yeager, "Mises and Hayek on Calculation and Knowledge," *Review of Austrian Economics* 7, no. 2 (1994): 93–109. Also, Israel Kirzner asserts that the Mises and Hayek contributions to the calculation debate, "are simply ways of expounding the same basic, Austrian insight, viz., that only market processes are able to harness the *discovery* potential of entrepreneurial competition." *Italics in the original*, see Israel Kirzner, "Book Review of *Hayek, Coordination and Evolution*," *Southern Economic Journal* 61, no. 4 (April 1995): 1244. If Kirzner is correct, it would seem that Mises and Hayek were both Kirznerians and the entire calculation debate was a debate about Kirzner's concept of entrepreneurship. On his concept of entrepreneurial discovery, see Israel Kirzner, *Competition and Entrepreneurship* (Chicago: University of Chicago Press, 1973) and *Discovery and the Capitalist Process* (Chicago: University of Chicago Press, 1985).

²Yeager, "Mises and Hayek," p. 94. *Italics in the original*.

The Review of Austrian Economics Vol. 9, No. 1 (1996): 151–62
ISSN 0889-3047

"it is wholly beside the point, because it rests on a gross misinterpretation of the meaning explicitly attached to the term 'calculation problem' by SRH."³ In response to Yeager, Salerno says,

it does not follow that, for SRH, *the* calculation problem as Mises conceived it refers narrowly to the mathematical techniques employed for manipulating the given quantitative data; it refers, instead, to the origination and meaningfulness of the data themselves. It is, in short, a problem of "appraisement" and not of "arithmetic."⁴

From this beginning point, he proceeds to cogently rebut Yeager's claim by demonstrating that entrepreneurial appraisal is not subsumed under market information.

Yet Yeager seems to imply something else in his claim that by its nature goes untouched by Salerno's rebuttal. Yeager seems to imply that the arithmetic facet of Mises's calculation argument is trivial. This claim is not only false but is odd coming from a student of Mises's work; for Mises made several true and non-trivial arguments based solely on arithmetic or mathematics and statistics, more generally: the impossibility of interpersonal utility comparisons (lack of a unit of subjective value), the impossibility of economic calculation (inability of comparing heterogeneous units of factors of production), the impossibility of mathematical equations in economic theorizing (lack of constants in human action), and the impossibility of statistical analysis in economic theory (lack of a probability density function for the data of human action).⁵ Acceptance of these *merely* arithmetic, mathematic, and statistical points destroys several major branches of orthodox economic theory: utility and welfare, socialist, mathematical, macroeconomics, and econometrics. Together these constitute a significant portion of what passes for economic thought today.

While it is true that Mises's calculation argument is not *merely* arithmetic; it is also true that it is not *merely* appraisement. Mises argued that economic calculation is a problem of both

³Joseph Salerno, "Reply to Leland B. Yeager on 'Mises and Hayek on Calculation and Knowledge,'" *Review of Austrian Economics* 7, no. 2 (1994): 112.

⁴*Ibid.* Italics in the original.

⁵Mises discusses each of these points in *Human Action: A Treatise on Economics* (Chicago: Henry Regnery, [1949] 1966).

arithmetic and appraisalment.⁶ More precisely, Mises's calculation argument has two dimensions: the impossibility of central planners performing the arithmetic of profit and loss computation in pure socialism which, in turn, makes it impossible for them to engage in entrepreneurial appraisals necessary to give meaning to profit and loss, and, thus, rationally allocate factors of production.⁷ Although information enters into the latter, it cannot enter into the former.⁸

The arithmetic facet of Mises's argument deals with the existence, or lack thereof, of a format in which information can be put and appraisals can be made. A format is necessary because the "raw data" required to answer relevant economic questions posed by the operation of a social process of exchange and division of labor are denominated in incommensurate units. Unless these units can be converted into a common standard, they cannot be compared; unless they can be compared the economic questions cannot be answered. As Mises said of one socialist scheme of economic calculation, "Calculation in kind is to be substituted for calculation in terms of money. This method is worthless. One cannot add or subtract numbers of different kinds (heterogeneous quantities)."⁹ The impossibility of comparing the number of apples to the number of oranges is an arithmetic problem; and a

⁶I am not asserting that Salerno fails to understand or appreciate the arithmetic facet of Mises's concept of calculation. He mentions it twice in his "Reply to Yeager," (pp. 112 and 120) and it is this point that Yeager himself notices in Salerno's work. My contention is only that proper recognition of this facet of Mises's argument also defeats the Yeager position.

⁷By the phrase, "impossibility of performing the arithmetic of profit and loss computation," we do not have in mind what Yeager seems to accuse us of meaning. As Salerno says, "the Misesian demonstration of the logical impossibility of socialism is not predicated on the central planners' incapacity to perform tasks that can conceivably be carried out by individual human minds," including adding and subtracting. See Salerno, "Reply to Yeager," p. 112. The arithmetic problem of calculation is not the inability to add common units together, it is the absence of such units. No "advances in supercomputers" can overcome the impossibility of adding together apples and oranges.

⁸These two steps correspond to the two conditions Mises claimed were necessary for calculation to take place: voluntary exchange of all goods including factors of higher order and the use in these exchanges of money. The first is necessary to bring higher-order capital goods under the orbit of the entrepreneurial "intellectual division of labor"; the second is necessary because without it, "It would not be possible to reduce all exchange-relationships to a common denominator." See Ludwig von Mises, *Economic Calculation in the Socialist Commonwealth* (Auburn, Ala.: Ludwig von Mises Institute, [1920] 1990), pp. 17–18.

⁹Ludwig von Mises, *Human Action*, p. 703.

fundamental, not trivial, problem of arithmetic. Without its solution, no arithmetic operations can be conducted at all.

The profit and loss calculation solves the arithmetic problem inherent in answering both economic questions posed by the operation of a social process of exchange and division of labor: what consumer goods should be produced and which combination of factors of production should be used to produce each consumer good. The arithmetic problem of the first question is the incommensurability of the subjective values of different individuals who participate in the social process of exchange and division of labor. There are two dimensions to the impossibility of making interpersonal comparisons of utility: no unit can be defined for preferences since they are subjective and even if units of subjective value existed for each person, they would not be comparable from one person to another.¹⁰

The solution to the problem of the incommensurability of the subjective values of individuals and the answer to the question of what consumer goods should be produced to satisfy them lies in the possibility of market prices denominated in money. Consumers demonstrate their preferences for some goods relative to others by purchasing and refusing to purchase. Since all preferences are demonstrated using the same standard, viz. money, the effects of action based on these preferences, viz. money prices, are commensurate, and, therefore, formatted for meaningful economic calculation.

Entrepreneurs then impute market value to each factor of production according to its marginal value product via their demand for the factors. Factors prices are then determined by the intensity of entrepreneurial demand relative to the opportunity cost placed on them by their owners. These prices make the different units of the factors commensurate and therefore, permit entrepreneurs to efficiently allocate factors across the production of consumer goods.¹¹

As a student of Mises's work, Yeager is surely familiar with his account of the relationship between the subjective values of

¹⁰As Mises said, "In an exchange economy the objective exchange value of commodities enters as the unit of economic calculation. This . . . renders it possible to base the calculation upon the valuations of all participants in trade. The subjective use value of each is not immediately comparable as a purely subjective phenomenon with the subjective use value of other men. It only becomes so in exchange value, which arises out of the interplay of the subjective valuations of all who take part in exchange." Mises, *Economic Calculation*, p. 12.

¹¹*Ibid.*, p. 23.

consumers and market prices as well as the impossibility of interpersonal utility comparisons. Even for those economists, few in number and among whom one should not expect to find Yeager, who disagree with the latter claim, it would seem strange for them to characterize the problem of interpersonal utility comparisons as anything but an arithmetic problem. You can only add or subtract items of like units. This fact is both arithmetic and non-trivial. An entire branch of economics (welfare economics) crashed to the ground on this point and another branch (utility economics) was completely revamped because of it.¹² The arithmetic dimension of Mises's calculation argument is based on the same arithmetic truth that makes interpersonal utility comparisons impossible; and recognition of this fact helps clarify and strengthen instead of, "caricature and trivialize," Mises's argument as Yeager claims.¹³

Mises understood that the question of what consumer goods should be produced can be answered by the central planners and therefore, is not a barrier to the establishment of a centrally-planned economy.¹⁴ The planners can do this by simply substituting their preferences for the unknowable and incomparable preferences of consumers. They produce, or attempt to produce, the goods they themselves value. This solution, however, is arbitrary with reference to the preferences of consumers. These, the central planners cannot know and even if they did they could not make the relevant comparisons to determine what subset of valuable goods should be produced to the exclusion of other goods consumers find valuable. Central planners with perfect information of consumer preferences still could not calculate what to produce to satisfy such preferences because they are ordinal rankings and therefore, cannot be compared. Even if central planners had perfect information of the subjective values of each individual denominated in units, they could not perform economic calculation because it is impossible to compare any items that are denominated in dissimilar units. Only if the central planners knew how to

¹²The old welfare economics and utility economics were based on the concept of cardinal utility which embodied two arithmetic mistakes: units of subjective value are possible and such units are interpersonally comparable. See Murray N. Rothbard, "Toward a Reconstruction of Utility and Welfare Economics," in *On Freedom and Free Enterprise*, Mary Sennholz, ed. (Princeton: Van Nostrand, 1956), pp. 224–62.

¹³Yeager, "Mises and Hayek," p. 94.

¹⁴Mises, *Human Action*, pp. 695–98.

convert the subjective units of each individual into a common standard would they be able to perform this part of economic calculation.

The central arithmetic facet of Mises's calculation critique is the incommensurability of the different factors of production that could be combined in different ways to produce each consumer good. Hours of labor cannot be compared to acres of land nor can these units be compared to units of each capital good. As Mises, discussing his example of central planners contemplating building a railroad, wrote in 1920, "Where one cannot express hours of labor, iron, coal, all kinds of building material, machines and other things necessary for the construction and upkeep of the railroad in a common unit it is not possible to make calculations at all. The drawing up of bills on an economic basis is only possible where all the goods concerned can be referred back to money."¹⁵ Nearly thirty years later, he wrote,

The director wants to build a house. Now, there are many methods that can be resorted to. . . . Which method should the director choose? He cannot reduce to a common denominator the items of various materials and various kinds of labor to be expended. Therefore he cannot compare them. . . . In short, he cannot, in comparing costs to be expended and gains to be earned, resort to any arithmetical operation.¹⁶

Concerning the pricing process of the market by which economic calculation solves the problem of incommensurability, Mises concluded that socialism cannot reduce the value of the means of production to "the uniform expression of a money price." In a market economy, "all prices can be referred back to a common expression in terms of money."¹⁷

If there were no arithmetic facet of this "common expression in terms of money," (contrary to Mises's explicit statement that there is) then the problem of economic calculation would not exist since the planners could discover the value of each factor in each use by withdrawing it.

Mises summed up the problem of calculation in socialism by saying, "In the main, socialist production might only appear

¹⁵Mises, *Economic Calculation*, p. 25.

¹⁶Mises, *Human Action*, p. 698.

¹⁷Mises, *Economic Calculation*, pp. 23-24.

rationally realizable, if it provided an objectively recognizable unit of value, which would permit of economic calculation in an economy where neither money nor exchange were present.”¹⁸ If this problem has no *merely* arithmetic facet, then why did socialists struggle to employ the labor theory of value to solve it? Mises finished the quote above by saying, “And only labor can conceivably be considered as such.” But, why not perform economic calculation in all factors of production at once claiming each of them to have intrinsic value and thereby dispense with the search for a “socially necessary” amount of labor, i.e., a common unit of labor in which all factors can be rendered? The existence of cardinal units is not sufficient for economic calculation to be performed. One cannot add together factors denominated in incomparable cardinal units, nor compare the efficiencies stated in cardinal numbers, e.g., the average product of labor with the average product of capital, of different factors of production. The task of economic calculation requires, in addition to cardinal units, a method by which the different units can be transformed into a common cardinal unit.¹⁹ If it is not necessary to have a common objective unit in which all factors can be meaningfully compared, then a large part of the debate about the labor theory of value was so much spilled ink.

Yeager’s contention about the arithmetic facet of Mises’s argument makes it neither erroneous nor trivial. To the contrary, it is both correct and devastating to naive socialists who believe that the economic problem of factor usage can be solved by central planners in the absence of profit and loss calculation based on monetary prices, i.e., in pure socialism, including those who think the problem could be solved by “advances in super-computers.”

It is only to defeat those socialists who wish to enter the debate on economic theory that Mises moves to more complex dimensions of his calculation argument.²⁰ To the assertion that socialism can overcome the incommensurability of different factors by having central planners set monetary prices for all goods and factors, Mises responds that the problem is calculation of objective value,

¹⁸Ibid., p. 33.

¹⁹These two issues, the existence of cardinal units and the existence of a common denominator into which the various cardinal units can be put, are analogous to the two dimensions of the impossibility of making interpersonal comparisons of utility discussed above.

²⁰Mises moves to some of these steps, in a different order than that presented here, when addressing a list of suggestions for socialist economic calculation in *Human Action*, pp. 703ff.

not objective units *per se*. Such a procedure would not solve the allocation problem since it leads to a “solution” that is arbitrary even from the viewpoint of the central planners, let alone that of consumers. The problem of factor usage cannot be solved by having the central planners assign a monetary wage to be multiplied by labor hours, and so on for each factor, so that the monetary costs of different combinations of factors capable of producing a given consumer good can be compared and the least cost method selected. Such cost calculations have no relationship to the preferences placed on the consumer goods and therefore, are useless for economic calculation. Only the market process can connect the value of factors to the value of consumer goods in a meaningful way.

Mises demonstrates this point by allowing that a socialist state could have a medium of exchange, limited in its scope to trading in some consumer goods. But, as he said,

where the means of production are state controlled . . . because no production good will ever become the object of exchange, it will be impossible to determine its money value. Money could never fill in a socialist state the role it fills in a competitive society in determining the value of production goods. Calculation in terms of money will here be impossible.²¹

To the assertion that the central planners can overcome the arbitrary nature of prices set by their own decree by having the managers of state-operated production facilities act as if they were entrepreneurs engaged in trade, Mises argues that one cannot “play” market.²² For entrepreneurial competition to perform the function of factor evaluation, the possibility of bearing the opportunity costs of different factor allocations must be real. Only with private property can entrepreneurs and capitalists risk their own wealth in the process of social production and therefore be in a position to make accurate appraisals of factor values.^{23, 24} To argue that play acting could mimic the results of

²¹Mises, *Economic Calculation*, p. 6.

²²Mises, *Human Action*, pp. 707–9.

²³Mises, *Economic Calculation*, p. 28.

²⁴One particular target Mises aimed at was the “market socialism” of Oskar Lange in his, “On the Economic Theory of Socialism,” reprinted in *On the Economic Theory of Socialism*, vol. 2, Benjamin Lippincott, ed. (Minneapolis: University of Minnesota Press, 1938), pp. 57–129. Yeager, in using the debate between Mises and Lange as the text for criticizing the SRH view, reveals the source of his lack of appreciation for the arithmetic facet of Mises’s calculation argument. See Yeager, “Mises and Hayek,” pp. 103ff. Mises had no need to mention the arithmetic

the market was to confuse the functions of management with those of entrepreneurship.

One cannot *play* speculation and investment. The speculators and investors expose their own wealth, their own destiny. This fact makes them responsible to the consumers. . . . If one relieves them of this responsibility one deprives them of their very character. They are no longer businessmen, but just the group of men to whom the director has handed over his main task, the supreme direction of the conduct of affairs. Then they—and not the nominal director—become the true directors and have to face the same problem the nominal director could not solve: the problem of economic calculation.²⁵

To the assertion that the central planners can overcome the “game-playing” nature of market socialism by using the pre-existing market set of prices, i.e., those prices existing in the capitalist system just prior to socialization, Mises argues that the transition from capitalism to socialism is too fundamental for the old prices to bridge the gap and that pricing must be “dynamic” since underlying economic phenomena are constantly changing. By destroying the differences in wealth in the existing market economy when expropriating private property, socialism disconnects the prices that correspond to those inequalities with the different conditions now prevailing for which calculations must be made. Moreover, any changes in conditions that underlie the economic allocation of factors makes the existing set of prices obsolete, and all the more so the greater the extent of such changes.²⁶

problem in response to Lange; market socialism overcomes that problem by employing money—the common denominator necessarily absent in a pure socialist system—and money prices. Mises was, thus, required to move to more complex dimensions of his calculation argument and criticize market socialism for its inability to perform entrepreneurial appraisals based on money prices which are not established in market exchanges of private property.

Because Rothbard fails to mention the arithmetic facet of calculation but does mention information in discussing the debate between Mises and Lange, Yeager attempts to construe Rothbard as once holding the Yeager position and then shifting to the SRH view. See Yeager, “Mises and Hayek,” p. 106. But Rothbard had no more reason to mention the arithmetic facet of calculation in this context than did Mises. Moreover, neither Mises, nor Salerno, nor Rothbard, nor I claim that the central planners do not face an information problem. The SRH claim is that Mises’s calculation argument has more to it than the information problem. Yeager’s claim that it does not is not proven by noting that Mises and SRH recognize information as a problem.

²⁵Ibid., p. 709. Italics in the original.

²⁶Mises, *Economic Calculation*, pp. 25–26.

Furthermore, as Salerno pointed out, Mises understood that answering the economic questions of what and how to produce requires entrepreneurs to correctly project appraisals of goods and factors into the future.²⁷ Since the data are continually changing, static modeling cannot be substituted for entrepreneurs to perform economic calculation. Comparative statics serves no better since it cannot determine how human action moves the solution from one point to another.²⁸

Moreover, general equilibrium is irrelevant to the actual problem that economic calculation must solve and that can be done so only by entrepreneurial activity. Neither the actual prices, both present and future, nor the preferences necessary for factor allocations to be made have any relationship to those of equilibrium. As Mises said, "what impels a man toward change and innovation is not the vision of equilibrium prices, but the anticipation of the height of the prices of a limited number of articles as they will prevail on the market on the date at which he plans to sell."²⁹

General equilibrium equations are formed by knowing the constants of those equations, under the assumption that no further change in the data is permissible. Without the assumption of no further changes, no constants exist and no equations can be formed. Yet, the economic system cannot achieve, or move toward, the equilibrium without changes from the existing set of data. The equations are, thus, useless for the task of allocating factors of production toward their general equilibrium uses. As Mises said, "What acting man needs to know is not the state of affairs under equilibrium, but information about the most appropriate method of transforming, by successive steps, [the total supply of produced factors allocated as they are today] into [the total supply of produced factors allocated as they need to be in equilibrium]. With regard to this task the equations are useless."³⁰

Even if the central planners had full knowledge of the state of general equilibrium and could see how to move production from original factors to the final equilibrium state, this would not suffice to circumvent the problem that only economic calculation can solve. The existing state of production does not correspond to any state of this perfect-knowledge production process. Existing

²⁷Salerno, "Reply to Yeager," pp. 120–23.

²⁸Mises, *Human Action*, pp. 710–11.

²⁹*Ibid.*, p. 711.

³⁰*Ibid.*, pp. 712–13.

capital goods embody past allocation errors relative to their perfect knowledge uses. Since these capital goods can neither be freely transferred into other uses nor transferred efficiently without taking account of their existing characteristics, central planners with perfect knowledge would still need to resort to economic calculation to properly allocate them. Mises concludes his discussion of economic calculation at this step where no recourse is made to the arithmetic facet of the argument when viewed in its entirety.³¹

Instead of realizing the logical construction of Mises's argument—beginning with its arithmetic facet and then in turn allowing, for the sake of argument, that the central planners can overcome progressively more difficult aspects of the calculation problem—Yeager implies that SRH assume that Mises was conceding that the central planners could solve these problems. Yeager says,

The necessary preparations for the vast central calculation, let alone the calculation itself, could *not* be accomplished; they are, to use Mises's word, "impossible." It seems perverse, then, to interpret Mises as nevertheless conceding the possibility of all those preparations and of balking only at the possibility of the calculation itself.³²

But Mises did not concede that a "preparation" or "information" problem could be solved by the central planners in the actual operation of socialism. He conceded the solution to these problems, *for the sake of argument*, for the very purpose of demonstrating that his calculation argument proved the impossibility of economic calculation, *even if these problems were solved*. The fact that he chose this method of argumentation is proof that his calculation argument has more to it than just the lack of information available to central planners.

In fact, Mises "concedes" much more than the solution to the "information" problem, in the final step of his argument. If Yeager has this perfect-information scenario in mind in his quote at the beginning of this article, then he misstates Mises's hypothetical conditions (under which there is no arithmetic facet of the argument). Mises is not, here, assuming that the central planner has

³¹Ibid., pp. 713–14.

³²Yeager, "Mises and Hayek," p. 101. Italics in the original. Also, see his other statements on Mises's concessions, pp. 97–98.

perfect information and *therefore*, can perform economic calculation, as Yeager implies in his quote. Mises is assuming that the central planner has “miraculously” solved the problems of economic calculation—not just information but calculation itself—and could therefore construct a perfect production structure over time, starting without any capital goods, to achieve some final equilibrium state. Even if the central planners had perfect information and the ability to calculate with that information, however, they still could not calculate how to effectively operate any actual existing economy they are attempting to control.

If Yeager means what he seems to say—that Mises could not have meant that a central planner with perfect information about preferences and factor conditions could not perform the arithmetic operations necessary to calculate—then he is wrong; for this is precisely the first step of Mises’s argument demonstrating the impossibility of economic calculation in the socialist commonwealth.³³

On the importance of the arithmetic aspect of the economic calculation, Mises said,

every action can make use of ordinal numbers. For the application of cardinal numbers and for the arithmetical computation based on them special conditions are required. These conditions emerged in the historical evolution of the contractual society. Thus the way was opened for computation and calculation in the planning of future action and in establishing the effects achieved by past action. Cardinal numbers and their use in arithmetical operations are also eternal and immutable categories of the human mind. But their applicability to premeditation and the recording of action depends on certain conditions which were not given in the early state of human affairs, which appeared only later, and which could possibly disappear again . . .

Modern civilization is above all characterized by the fact that it has elaborated a method which makes the use of arithmetic possible in a broad field of activities. This is what people have in mind when attributing to it the—not very expedient and often misleading—epithet of rationality.³⁴

³³Mises explicitly made these assumptions in the development of his calculation argument. In addition to the statements already quoted, see Mises, *Human Action*, p. 696.

³⁴*Ibid.*, p. 199.

Review Essays

Keynes Was a Keynesian

Roger W. Garrison

The State of Interpretation of Keynes

Edited by John B. Davis

Boston: Kluwer Academic Publishers, 1994

A half dozen papers together with formal comments and an introduction have been assembled to help establish the state of interpretation of Keynes. The contributors to this volume are ideologically like-minded but geographically diverse (Australia, Brazil, Canada, Great Britain, Italy, and the United States are represented). Their book, whose title belies a certain narrowness of focus, is not for everyone; it reflects the concerns of one particular Keynesian school, best described, in my judgment, as Post Keynesian fundamentalism.

Academics who continue to be amused and intrigued with the still-growing literature on the economics of John Maynard Keynes have had to learn to distinguish among the several different schools that draw inspiration from the master. Hyphenated or adjectival Keynesianism includes, for instance, both Neo-Keynesianism, which is based on an assumed wage and price stickiness, and New Keynesianism which attempts to explain the stickiness. Neo- and New Keynesianism share certain methodological presuppositions with Neo- and New Classicism but do not share in the judgment that markets are generally self-equilibrating. Interpreters who prefer to blend Keynes's ideas with those of the old classical school, which featured a cost-based production theory, have adopted "post" as their adjective of choice. Readers of this literature have been asked to maintain a distinction between Post Keynesianism and Post-Keynesianism—a distinction as subtle as

Roger W. Garrison is associate professor of economics at Auburn University, Auburn, Alabama.

The Review of Austrian Economics Vol. 9, No. 1 (1996): 165–71
ISSN 0889-3047

the difference in labeling.¹ But whatever the particular labeling, most interpreters have come to see as virtual opposites *Keynesian Economics* and *the Economics of Keynes*, as contrasted in the title of Axel Leijonhufvud's 1968 book.

Murray Rothbard used to proclaim gleefully that "Keynes was a Keynesian."² He took great pleasure in the irony that by so proclaiming he set himself apart from most all modern interpreters. Keynes favored monetary manipulation and fiscal activism, deficit finance and income redistribution, all for the purpose of spending our way out of depression. When his attention turned from short-run policy to long-run reform, his enthusiasm for these stop-gap measures gave way to his anticipations of a future utopia—and to schemes for ensuring and hastening its arrival. The inherent uncertainty of the future, in his view, gave centralized decision making a clear advantage over the decentralization that characterizes market economies. Keynes advocated the "socialization of investment" and the "euthanasia of the rentier." The rate of interest, which "rewards no genuine sacrifice," could and should be driven to zero, at which point capital would cease to be scarce and the distribution of income would be more equitable. In a matter of two generations, the economic problem of scarcity can be solved, such that our grandchildren can occupy themselves with questions of aesthetics rather than questions of economics.³ This is the uninterpreted Keynes.

Post Keynesians emphasize Keynes's vision of utopia and the associated reform proposals almost to the exclusion of his diagnosis of depression and prescription of short-run demand-management

¹Paul Davidson defines the unhyphenated Post Keynesians very broadly and then identifies a classically oriented subset, to which he awards the hyphen. Further, he prefers to call Post-Keynesians Neo-Keynesians, a label I have used for the Keynesian component of the Keynesian-Neoclassical Synthesis found in modern textbooks. See Davidson, "Post Keynesian Economics: Solving the Crisis in Economic Theory" in Daniel Bell and Irving Kristol, eds., *The Crisis in Economic Theory* (New York: Basic Books, 1981), pp. 151–73.

²That Keynes was a Keynesian—of the much derided Keynesian system provided by Hicks, Hansen, Samuelson, and Modigliani—is the only explanation that makes any sense of Keynesian economics." Murray N. Rothbard, "Keynes, the Man" in Mark Skousen, ed., *Dissent on Keynes: A Critical Appraisal of Keynesian Economics* (New York: Praeger Publishers, 1992), p. 196.

³Quoted phrases are from John Maynard Keynes, *The General Theory of Employment, Interest, and Money* (London: Macmillan, 1936), p. 376–78. For assessments of the utopian Keynes, see Joseph T. Salerno, "The Development of Keynes's Economics: From Marshall to Millennialism," *Review of Austrian Economics* 6, no. 1, (1992): 3–64, and Roger W. Garrison, "Keynesian Splenetics: From Social Philosophy to Macroeconomics," *Critical Review* 6, no. 4 (1993): 471–92.

policies. In fact, standard textbook Keynesianism, whose graphs and equations make the case for monetary and fiscal activism, are repeatedly described in the Davis volume as “bastardized Keynesianism” (Joan Robinson’s term) so as to provide an appropriate contrast with the more radical Keynesianism adopted by the volume’s contributors. If Post Keynesians did nothing but embrace these utopian aspects of Keynes, they would more accurately be described as Keynesian fundamentalists. But they do more. They add to the chronic demand deficiencies featured in Keynes’s *General Theory* the ideas about supply first articulated by the classical economists and subsequently exhumed by Piero Sraffa just before the Keynesian Revolution. Sraffa’s *Production of Commodities by Means of Commodities*, written during the 1930s though not published until 1960, is offered as the supply-side counterpart to Keynes’s demand-side theorizing. Michael Lawlor’s contribution, “The Own-Rates Framework as an Interpretation of the *General Theory*: A Suggestion for Complicating the Keynesian Theory of Money,” identifies Sraffa’s (ill-tempered) review of Hayek’s *Prices and Production* as the foundation underlying this new approach to monetary theory adopted by Keynes. As a duo, however, Sraffa and Keynes give us an unlikely and ill-fitting rendition of supply and demand.

Post Keynesians typically—but not in the volume under review—suggest still more complications: a dual market structure consisting of both competitive and oligopolistic firms, mark-up pricing (practiced by the oligopolists) to finance new investment, and Marxian class conflict. The Davis volume avoids what would otherwise appear to be a hopeless grab bag of ideas by focusing on Keynes—his vision, philosophy, methods, analyses, and tactics. One of the formal comments (Allin Cottrell’s) even includes a warning against perceiving Keynes as “a closet post Keynesian, harboring all kinds of heresies but suppressing them for the sake of perceived polemical advantage!” Another comment (Robert Prasch’s) questions the idea that Sraffian supply fits well with Keynesian demand. Had Keynes thought it did fit well, why did he not say so? Why did he say, instead (in his often quoted letter to George Bernard Shaw), that “when my book has been duly assimilated and mixed with politics and feelings and passions, . . . the Ricardian foundations of Marxism will be knocked away”? Prasch is persuasive that what Keynes believed would be “knocked away,” here, was Sraffa. If the grab-bag features of Post Keynesianism are left undiscussed while the relevance of Sraffa is in some doubt, then the term Post Keynesian

fundamentalism seems appropriate. Further, the volume's editor, if not all the contributors, may not be so tightly bound to the "Post," in which case we would be back to the uninterpreted utopian Keynes.

It is one thing to proclaim that Keynes was a Keynesian, (again) as Rothbard so often did; it is quite another to treat Keynes's vision as a relevant or fruitful reflection of economic reality. But *The State of Interpretation of Keynes* will strike the reader as something of a stocktaking; the papers constitute background reports that form the basis for updating the Post Keynesians' research agenda. As explained in the editor's introduction, Keynes's ideas deserve—and now may get—a fresh hearing. The particular occasion for this stocktaking together with the particular focus of several of the papers is both curious and revealing.

There are two circumstances that have combined to put the original Keynes back in the limelight. First, there is the crumbling of bastardized Keynesianism with its graphical apparatus and algebraic expression of government-spending multipliers and tax multipliers and all that. Neoclassical modes of thought have long restricted the thinking of mainstream economists and have made Keynesian economics look very unKeynesian. Recent trends, according to Davis, have been in the direction of flouting the strictures imposed by formal Neoclassicism and exploring a number of issues, including those of power and entrepreneurship. Methodological permissiveness may allow economists to look beyond the formalized income-expenditure relationships and get a fresh view of the *General Theory*. Second, Keynes's ideas seem to Davis to have a certain relevance in Keynes's time and in ours—a relevance that may have been lacking during most of the post World War II period. The decades-long capitalist-communist stalemate separated two periods characterized by political pluralism, relative disorder, and internationally based uncertainties. So the time is right, once again, to focus on the expectations that govern the economic process in the face of uncertainty and to think about the institutional arrangements most conducive to a healthy economy.

Davis finds similarities in the 1930s and the 1990s in terms of global political turbulence. He fails, however, to give due weight to the key difference between these two periods. In the 1930s the world was moving rapidly in the direction of centralization; in the 1990s it is undergoing substantial decentralization. Ironically, the perceived similarities of circumstances then and now cause Davis and several other contributors to this volume to rethink

the meaning of Keynes's call for "a somewhat comprehensive socialization of investment."⁴ A healthier reading of world history might have caused these scholars to rethink the meaning of Ludwig von Mises's claim of the "impossibility of economic calculation under socialism."⁵ In fact, the very concepts that Davis hopes to see back in play, such as the concepts of power and entrepreneurship, help to bolster Mises's case for decentralized decision making.

As with almost every other aspect of Keynes's writing, the phrase "socialization of investment" is in for some interpreting. What did Keynes have in mind? While no one believes that he was thinking of outright state ownership of the means of production, other possible meanings involve further questions that neither Keynes nor his followers have adequately addressed. It is clear in his discussion following the call for socialized investment that Keynes is concerned with the "volume" and not the "direction" of employment. Keynes argues as if the government—or rather, "forces outside the classical scheme of thought"⁶—could control the volume without affecting any other aspect of the market economy. What sort of powers would government have to wield to be able to exert such a force? And how would the quality of entrepreneurial decisions be affected if entrepreneurs had to anticipate the use—and possible misuse—of such powers? There are no answers to these questions that put socialization in a favorable light. The simple fact is that the conceptually distinct aspects of "volume" and "direction" as applied to employment or output are governed by a single set of market forces. Joan Robinson, who recognized the actual unity of these market forces but favored a more wholesale form of socialization, chided Keynes for even wanting to control volume without controlling direction. Direction, in her view, needed some controlling, too.⁷

In an alternative interpretation, one discussed in Hans Jensen's contribution, "forces outside the classical scheme" are not exerted by the state *per se* but rather by semi-public bodies. Keynes seemed to have envisioned large, privately owned firms

⁴Keynes, *General Theory*, p. 378.

⁵Ludwig von Mises, *Human Action: A Treatise on Economics*, 3rd rev. ed. (Chicago: Henry Regnery Company, 1966), p. 698.

⁶Keynes, *General Theory*, p. 378.

⁷Joan Robinson, "What Has Become of the Keynesian Revolution?" in Milo Keynes, ed., *Essays on John Maynard Keynes* (Cambridge: Cambridge University Press, 1975), pp. 123–31.

with public-spirited managers. Readers of Keynes or of the Davis volume can only wonder what kind of power the state needs in order to maintain a public-spiritedness among these managers—and, again, what this kind of power might do to the entrepreneurial spirit among them and others. A discussion by Robert Dimand offers still another interpretation of the socialization of investment. Several years before the *General Theory*, Keynes had proposed a National Investment Board that would give the government some control over the volume of investment. An NIB, though, sounds suspiciously like a peacetime version of the United States' War Industries Board of the World War I era. Rothbard is noted for identifying the WIB as an important precursor of the subsequent New Deal policies that came to be closely identified with so-called bastardized Keynesianism.⁸

The Post Keynesians seem to be seriously concerned about what, exactly, Keynes may have had in mind. Many of the contributors, and especially Davis himself in his discussion of "Keynes's Philosophical Thinking," employ a what-did-he-know-and-when-did-he-know-it strategy in their attempts to sort things out. Davis also warns against a careless mixing of Keynes's philosophy and economics: Ideas are not easily transplanted whole across decades and disciplines. The alternative that none of the contributors considered is the simplest one—though not the most satisfying for those who consider Keynes the fountainhead of economic wisdom: Keynes himself did not know what he meant. Keynes did not know what the appropriate supply-side counterpart to his theory was. He did not think through the implications—or recognize the virtual impossibility—of a zero rate of interest. He did not know how, exactly, a comprehensive socialization of investment could be implemented. Neither, we might add, did the more radical socialists have any clear ideas on the particulars of the economic mechanisms in a socialist state. These and similar answers to many other questions seem to be the most plausible basis for understanding the *General Theory*.

The viability of Keynesian economics, bastardized or otherwise, may be as hotly contested now as it was in Keynes's own time. And the Post Keynesian interpreters, with or without the

⁸Murray N. Rothbard, "War Collectivism of World War I," in Ronald Radosh and Murray N. Rothbard, eds., *A New History of Leviathan: Essays on the Rise of the American Corporate State* (New York: E. P. Dutton, 1972), pp. 66–110.

hyphen, may never discover just what, exactly, Keynes meant. But as interpretations continue to proliferate, this collection of papers provides its readers a close encounter with the original Keynes. The experience may cause them to doubt whether Keynesianism in any of its guises can provide a healthy understanding of economic reality or a suitable guide to prescribing policy or proposing reform.

Intimidation by Rhetoric

Murray N. Rothbard[†]

Knowledge and Persuasion in Economics

Donald N. McCloskey

Cambridge: Cambridge University Press, 1994

Professor Donald McCloskey of the University of Iowa in recent years has spectacularly “reinvented” his career. At first, he built a solid reputation as a journeyman economic historian. Later, he cleverly used that base to make a histrionic Great Leap Skyward, to become one of economics’ best known gadflies, critics, and methodologists. In particular, McCloskey has invented a new methodology called “Rhetoric,” an import from literary theory. In his latest book (p. xv), his third on rhetoric and economics, he “sets economics beside philosophy,” turning his attention to philosophy, and attempting to refute his myriad of critics within the economics profession.

“Rhetoric” is kissin’ cousin to the short-lived “hermeneutics” movement in economics, and indeed he helped contribute to that movement.¹ Like his friends the hermeneuticians, McCloskey insists that economics must be an eternal and perpetually “open” “conversation.” Like the hermeneuticians, he is a nihilist and a

Murray N. Rothbard (1926–1995) was S. J. Hall distinguished professor of economics at the University of Nevada, Las Vegas and editor of *The Review of Austrian Economics* (1987–1995).

¹Donald N. McCloskey, “Storytelling in Economics,” in Don C. Lavoie, *Economics and Hermeneutics* (London and New York: Routledge & Kegan Paul, 1990), pp. 61–75. For a critique of hermeneutics, see David Gordon, *Hermeneutics vs. Austrian Economics* (Auburn, Ala.: Ludwig von Mises Institute, 1986); Murray N. Rothbard, “The Hermeneutical Invasion of Philosophy and Economics,” *Review of Austrian Economics* 3 (1989): 45–59; and Hans-Hermann Hoppe, “In Defense of Extreme Rationalism: Thoughts on Donald McCloskey’s *The Rhetoric of Economics*,” *Review of Austrian Economics* 3 (1989): 179–214.

The Review of Austrian Economics Vol. 9, No. 1 (1996): 173–78
ISSN 0889-3047

village relativist. The difference is that McCloskey is far shrewder than they. In two ways: first, because he has a lock on the term "rhetoric," and can ring the changes on that word; and second, because he is far more elusive and evasive, far more the "artful dodger," than his stodgy and plodding hermeneutical comrades.

Rhetoric, McCloskey tells us, is the art of persuasion, and McCloskey has certainly learned the dark side of this art. The first phase of his rhetorical game was to trumpet himself as a great writer, ridiculing various established economists for terrible, jargon-filled writing. Here, McCloskey was shooting fish in a barrel; most economists have horrendous writing styles, *and they know it*. Hence, McCloskey was hitting mainstream economists where they were most vulnerable. As a result, McCloskey was able to establish himself in the minds of the profession as a great stylist and writer. Using rhetoric as intimidation, he was able to bully the rest of the profession into handing him, without real evidence or analysis, this crucial accolade.

In actual fact, McCloskey is not even a good, much less a great, writer. His new book, for example, is an inchoate mishmash of assertion, gossip, anecdote, phony Socratic dialogue, off-hand attack, and unsupported rumor. Gossip, of course, is easier, and certainly more fun, to read than theoretical argument, and McCloskey has now established himself as the Cindy Adams of the economics profession, methodological branch. Indeed, it is curious that a distinguished university press should lend its imprint to this farrago, presumably because they too were intimidated by McCloskey's self-promotion as a great stylist. Actually, two truly fine writers are both *bêtes noires* of McCloskey: Stigler and Schumpeter. Both were all that McCloskey is not: lucid, penetrating, and gifted with a graceful style and witty turn of phrase. But presumably, they were too old-fashioned and "linear" for our rhetorician.

Having once established himself as a great writer, McCloskey now could wield a powerful if implicit club over his critics. For it turns out, monotonously and one after the other, that none of McCloskey's critics understands him. Every one of them is mired in his preconceptions, in some cases his mind "hard wired" to reject the McCloskeyan light. Either they don't understand philosophy, or they understand philosophy all too well. If they are neoclassicals or various brands of positivists, they are blind to McCloskeyan insights. If they think that "rhetoric" or "metaphor" is unscientific, then don't they know that all science is

“metaphor” anyway, that presumably there is no difference between the law of gravitation and one of McCloskey’s gossipy pronouncements.

But McCloskey is nothing if not an evasive and slippery brokenfield runner, always presenting a moving target. To those neoclassical economists who attack him as a nihilist and an enemy, he protests that, no no, he too is down deep a neoclassical, except a neoclassical on a higher and more sensible methodological level. To the leftists and institutionalists who embrace McCloskey for the same reason, he protests similarly that while *his* critiques of mainstream economics seem to be the same as theirs, he is really a mainstreamer. One by one, none of his critics get it; everyone is blind, confused, out of it.

Ordinarily, if we find a writer, in whatever discipline, whom no one seems to understand, the onus will quickly begin to fall upon the writer himself: maybe no one understands him because he is incoherent, self-contradictory, in short, incomprehensible. But McCloskey had already cunningly precluded that reaction by entrenching himself as a uniquely great writer, so that the fault must always lie with his critics.

Throughout all of McCloskey’s razzle-dazzle, however, he really only has one arrow in his rhetorical bow. To McCloskey, the good is to be “open,” the bad to be “closed,” a viewpoint that he embraces as *Sprachethik*, a typical McCloskey ploy—making a term more portentous by putting it into a foreign language. And sure enough, at bottom McCloskey has only one rebuttal to his host of critics; *they*, without exception, are “closed,” whereas *he*, of course and by definition, is “open.” Q.E.D. And it is not very long into his discourse that we come to realize what, for McCloskey, is the practical definition of these two words: someone is “open” if they agree with him, “closed” to the extent that they disagree.

In short, in addition to being its Cindy Adams, Donald McCloskey is even more the Humpty-Dumpty of the economics profession. “When I use a word,” said this distinguished rhetorician who unfortunately had the satirical bent that the *serioso* McCloskey totally lacks, “it means just what I choose it to mean—neither more nor less.” To which the egg-shaped one significantly added: “The question is which is it to be master—that’s all.” Unfortunately, McCloskey has gotten away with his rhetoric of mastery far too often.

If we come down from the lofty heights of methodology, how accurate is McCloskey as an observer or historian of modern economic thought and opinion? Let us consider two examples. One

is his treatment of the allegedly "closed Austrianism" of myself and of *The Review of Austrian Economics* [RAE]. In this case we find the rhetorician as purveyor of unsubstantiated rumor: "The *Review* is known in Austrian circles as carefully excluding all but Rothbard's views" (McCloskey, *Knowledge and Persuasion*, p. 314). How does he establish this statement? McCloskey's one piece of "evidence" is that Hans-Hermann Hoppe, reviewer of McCloskey's *The Rhetoric of Economics* in the RAE, and not coincidentally one of its most effective critics, "is an academic colleague of Rothbard's," that is, at the University of Nevada, Las Vegas. Now, surely, McCloskey is familiar enough with academic life not to be employing this statement as anything but rhetorical deceit. Surely, he must know that to be in the same economics department as someone else does not imply agreement. Indeed, of the dozen and a half members of the UNLV economics department, Hoppe and I are the only ones whom any reasonable person would dub as an "Austrian" in any sense, whether "open" or "closed." McCloskey's only other argument is characteristic: the weasel phrase "known in Austrian circles." McCloskey, as a supposed scholar and historian, might have checked the record of the RAE before treating gossip as fact. If he had done so, he would have found that the RAE has included articles by such distinguished economists as Leland Yeager, Richard Timberlake, Lowell Gallaway, Richard Vedder, Charles W. Baird, Robert Higgs, Israel M. Kirzner, Bruce Caldwell, W.H. Hutt, and Gordon Tullock, not one of whom could be called, by even the most fevered imagination, a "closed Austrian" à la the dread Rothbard: "careful exclusion"?

The second example of his accuracy is the embarrassingly ignorant and self-contradictory way in which McCloskey treats the great Joseph Schumpeter. First, he claims Schumpeter as a paradigmatic neoclassical upholder of the Whig Theory of the history of economic thought, that is, a believer that any dominant economic thought is superior to that at any earlier period of time, since the history of thought consists in testing one's theories, improving one's logic, and marching step-by-step, over time, into the light of more and more complete truth.

There is indeed a grain of truth in pinning this label on Schumpeter (he certainly believed in the progress of economic thought, and he revered Walras, the founder of the neoclassical paradigm, as "the greatest economist of all time"), but to dub him a Whig historian can only be called a grotesque caricature. Later in the book (p. 146), McCloskey mentions Schumpeter's famous

attack on “the Ricardian Vice,” without realizing that one of the basic themes of Schumpeter’s famed *History of Economic Analysis* was a bitter and sustained attack upon Smith and Ricardo, their scuttling of the Scholastic and Continental tradition of scarcity-and-utility analysis of value, and their consequent shunting of economics onto a path (“classical economics”) from which it took economics a century to recover (in the form of Walras and the other “marginalists”). Whatever it is, a belief in a century-long and profound worsening of dominant economics can scarcely be called a “Whig theory”!²

Trying to squeeze Schumpeter into the positivist Stigler–Blaug mould, McCloskey then gets the “Ricardian Vice” all wrong, claiming that for Schumpeter, this vice is “applying blackboard propositions, untested, to the world” (McCloskey, *Knowledge and Persuasion*, p. 146). McCloskey cites page 473 of Schumpeter’s volume, but if we turn to that page we find a completely different account (as how could we *not* from a fervent adherent of Walras’s surely “untested” set of general equilibrium equations). For Schumpeter’s criticism is not at all that Ricardo’s theory was not empirically tested. Schumpeter denounces Ricardo, because, unlike his Continental predecessors and indeed unlike Walras to come, he was in no sense concerned with a “comprehensive vision of the universal interdependence of all the elements of the economic system.” Instead, Ricardo

cut that general system to pieces, bundled up as large parts of it as possible, and put them in cold storage—so that as many things as possible should be frozen and “given.” He then piled one simplifying assumption upon another until, having really settled everything by these assumptions, he was left with only a few aggregative variables between which, given these assumptions, he set up simple one-way relations so that, in the end, the desired results emerged almost as tautologies. For example, a famous Ricardian theory is that profits “depend upon” the price of wheat. And under his implicit assumptions . . . this is not only true, but undeniably, even trivially so. Profits could not possibly depend upon anything else, since everything else is “given,” that is, frozen . . . The habit of applying results of this

²Indeed, Schumpeter’s *History of Economic Analysis* (New York: New York University Press, 1954) was notable in its pioneering resurrection of Scholastic economics from its long misrepresentation and neglect at the hands of the dominant British and American tradition—scarcely the work of a Whig historian!

character to the solution of practical problems we shall call the Ricardian Vice.³

In the past two decades, the collapse of arrogant positivism in the philosophy profession, and the severe decline of the allied Keynesian–Walrasian neoclassical paradigm in economics, has led to a veritable “crisis situation” in all areas of economic theory, in particular among the methodologists who are concerned about the philosophical and methodological groundwork of economics. This crisis situation has been a healthy one, for it has mainly dislodged the previously ruling and false paradigms, and allowed room for the jostling of many new doctrines, competing for attention and dominance. But the flip side of this salubrious situation is that standards in general are lowered, and a lot of nonsense and absurdity gets to pass as sober and profound contributions to the discipline. Donald McCloskey and his inchoate and ill-informed “rhetoric” is an outstanding example. It is precisely in exciting times of methodological crisis that we need to have our built-in claptrap-detectors honed and sharpened. To the extent that they are, McCloskey’s work will receive the clear-eyed treatment it so richly deserves.

³Ibid., pp. 472–73.

Book Review

Economic Thought Before Adam Smith (vol. I)
and *Classical Economics* (vol. II)

Murray N. Rothbard

Aldershot, England and Brookfield, Vt.: Edward Elgar, 1995

On first seeing Murray Rothbard's two volumes, I regretted having agreed to review them, for they cover the history of thought only up to about 1870, when, in my view, economics begins to get really interesting. Actually reading them changed my mind. Rothbard—not just the libertarian guru but the joyful and indefatigable scholar—makes the thinkers even of reputedly dreary epochs come alive.

As he says, his history “is much longer than most since it insists on bringing in all the ‘lesser’ figures and their interactions as well as emphasizing the importance of their religious and social philosophies as well as their narrower strictly ‘economic’ views” (I, p. xiii). Lao Tzu (around sixth century B.C.), leader of “the world’s first libertarians,” receives respectful attention (I, pp. 23–26). So do the pre-Socratics, Xenophon, and, with much qualification, Aristotle, whereas Plato is classified as a “right-wing collectivist” (I, chap. 1). Even Amos Kendall gets credit for his subjective value theory and “the first expression of the law of diminishing marginal utility” in his Kentucky newspaper in 1820 (II, pp. 130–32). Rothbard regrets that Ricardo and his epigones did not read and understand Kendall’s rejection of any objective standard of value.

Appropriately in volumes subtitled *An Austrian Perspective on the History of Economic Thought*, Rothbard grades earlier economists high or low according to how they anticipated subjectivist and marginalist achievements or obstructed intellectual progress. He writes smoothly and, especially in his appraisals, with touches of humor.

Leland B. Yeager is professor emeritus of economics at Auburn University.

The Review of Austrian Economics Vol. 9, No. 1 (1996): 181–88
ISSN 0889-3047

His appraisals extend to earlier historians of thought. S. Leon Levy's biography of Nassau Senior is "chatty and uncomprehending" (II, p. 485). Samuel Hollander's "massive and bizarre project to transform all the classical economists into perfect little propounders of neoclassical, general equilibrium doctrine" suffered a "devastating and most welcome demolition" at the hands of T. W. Hutchison (II, p. 492).

Mostly, of course, Rothbard appraises the economists themselves. He lambastes David Ricardo's "abysmal writing, in style and organization" and goes on to quote Alexander Gray's similar judgment. "For all too many people," however, "obscurity and bad writing equal profundity." Obscurity has charms both for the great man and for the adepts who cluster around him (II, p. 103). With evident relish Rothbard quotes H. L. Mencken at length on the flatulent banalities of Thorstein Veblen's prose (II, p. 153).

Bishop Berkeley wrote *The Querist*, containing his major pronouncements on economic questions, "solely as a series of 900 loaded questions, by which [he] hoped to influence public opinion through sheer rhetoric without having to engage in reasoning" (I, pp. 331-32). "On his farm [J. C. L. Simonde de Sismondi] fought against overproduction in his own dotty way: making sure that production would be as low as possible by choosing the feeblest workers for employment on his farm, and deliberately having his house repaired by an incompetent worker" (II, p. 34).

"[Jeremy] Bentham functioned as the Great Man, scribbling chaotically on endless and prolix manuscripts elaborating on his projected reforms and law codes. . . . The affluent Bentham lived in a capacious house surrounded by flunkies and disciples, who copied revision after revision of his illegible prose to get ready for eventual publication. He conversed with his disciples in the same made-up jargon with which he peppered his writings" (II, p. 49). "[I]n 1804, Jeremy Bentham lost interest in economics, a fact for which we must be forever grateful. It is only unfortunate that this waning of zeal had not occurred a half-decade before" (II, p. 55). (Rothbard reinforces my own impression of Bentham as an unattractive person, illustrated by his obsession with his Panopticon project. I doubt, though, that Bentham's personality should automatically discredit all philosophical doctrines linked, however loosely, with his name.)

As for John Stuart Mill, "It is difficult to think of anyone in the history of thought who has been more egregiously and

systematically overestimated" (II, p. 491). Despite Gertrude Himmelfarb's notion of two Mills, the (good) conservative moralist and the (bad) libertarian, Rothbard sees "only one Mill—multi-faceted, self-contradictory, kaleidic, devious, muddled and filiopietistic" (II, p. 493). "It is impossible to estimate how much of John Stuart Mill's inveterate and eternal contradictions, qualifications and alterations were due to honest muddle-headedness and how much to devious and evasive intellectual broken-field running" (II, p. 279). (While not challenging Rothbard's judgments on specific points, I'll confess to liking much in Mill's work, particularly parts of *Utilitarianism*, *On Liberty*, and *Considerations on Representative Government*. Mill was not as pervasively wrong or evil as, say, Karl Marx. A reader may as well get what he can from a serious writer. When one is trying to think one's own way through some topic by pulling together one's own and other people's ideas, the question of *overall* assessment of those people, however unfavorable, need not intrude.)

Rothbard crisply summarizes his judgments on a couple of Marxian concepts. "Alienation', to Marx, bears no relation to the fashionable prattle of late twentieth century Marxoid intellectuals" (II, p. 349). Regarding the material dialectic, "It is difficult to state this position without rejecting it immediately as drivel" (II, p. 377).

As I have already suggested, Rothbard breaks away from the second-handism and parroting that characterizes much history of thought. He doubts the sainthood of Adam Smith (esp. I, chaps. 16 and 17). Smith slipped backward from David Hume's insights into monetary theory and balance-of-payments analysis. He dropped earlier contributions about subjective value, entrepreneurship, and emphasis on real-world markets and pricing and replaced it all "with a labor theory of value and a dominant focus on the unchanging long-run 'natural price' equilibrium, a world where entrepreneurship was assumed out of existence" (I, p. xi; cf. summary judgment at I, p. 501). He mixed up Calvinism with economics, as in supporting usury prohibition and distinguishing between productive and unproductive occupations. He lapsed from the *laissez faire* of several eighteenth-century French and Italian economists, introducing many waffles and qualifications. His work was unsystematic and plagued by contradictions. He came close to plagiarism while accusing others of it. Rothbard credits Paul Douglas for a relatively clear-eyed assessment at a University of Chicago commemoration

of the 150th anniversary of *The Wealth of Nations*. He also cites Joseph Schumpeter's dissent from the conventional admiration of Smith.

David Ricardo also fares badly at Rothbard's (and Schumpeter's) hands (II, chaps. 3 and 4). Ricardo dealt more in aggregates and in supposed long-run equilibria and other abstractions than in realities; he did not fully understand the principle of comparative advantage, which was less his own idea, anyway, than James Mill's; he was less interested in what fosters and what impedes the creation of wealth than in its distribution among broad social classes; he clung to a cost theory and indeed mostly a labor theory of value; he gave ammunition to Marx.

Reading Rothbard makes me more willing than before to confess my own judgments (derived from reading *The Wealth of Nations* and *The Principles of Economics and Taxation* longer ago than I care to admit) that Smith and Ricardo are downright tedious, especially taken in large doses.

Rothbard's adverse assessments are perhaps more fun to read and quote than the reverse, but I must not leave the impression that he gives out little praise. His heroes include several of the scholastics who flourished from the thirteenth century into the seventeenth century, including members of the Spanish School of Salamanca. They had insights into subjective-value theory and understood the "just price" as the competitive market price rather than as a theological concept. Richard Cantillon was "the founding father of modern economics," writing the first systematic treatise. A. R. J. Turgot made brilliant contributions in several areas, even though his writings on economics add up to fewer than 200 pages. J. B. Say had sound insights into value theory, macroeconomics, the role of the entrepreneur, the rationale of *laissez faire*, and methodology. Nineteenth-century debates on money and banking were instructive.

Anyone writing history, including intellectual history, must focus attention selectively. Rothbard's own preferences come across, but without unpleasant obtrusiveness. While not trumpeting it, he exudes an evident sympathy for workers, peasants, and the poor; he is no apologist for the rich and powerful. Repeatedly if briefly he shows sympathy with doctrines of natural law and natural rights and, more broadly (and regretably, in my own view), with some sort of anti-utilitarian ethics. He takes a Lockean position on the origins of legitimate property rights, correctly noting that a labor theory of property is not at all the same as a labor theory of value (I, pp. 56–58).

Rather to my surprise, he manages to find several predecessors for his own advocacy of 100-percent-reserve money (II, 210–16, chap. 14, and *passim*).

On a few points, Rothbard seems to have changed his mind, or his rhetoric. He does not insist as emphatically as before on a purely subjective theory of value and a pure-time-preference theory of interest. He explains that degrees of scarcity interact with subjective appraisals to determine marginal valuations and that objective factors help determine degrees of scarcity. In interest theory he illuminates the interaction of subjective time preference with the greater productivity of well-chosen roundabout methods of production (II, pp. 22–24, 139, 141); and he has words of praise for Böhm-Bawerk, who recognized this interaction of subjective and productivity factors.

In contrast with his own *America's Great Depression* (esp. chap. 4), where he used an imaginatively broadened definition of the money supply to discuss monetary inflation in the 1920s, Rothbard now reverts to the narrow medium-of-exchange definition (II, pp. 164, 183, 251). In scattered remarks on business-cycle theory, he retains a monetary interpretation without always insisting on the specific "Austrian" scenario of Mises and the early Hayek.

Rothbard no longer endorses Hans-Hermann Hoppe's claim to derive libertarian policy positions purely from the circumstances of discussion itself, without any appeal to value judgments (cf. *Liberty*, September and November 1988). On the contrary, and as he had done earlier, Rothbard now correctly observes that policy recommendations and decisions presuppose value judgments as well as positive analysis (II, p. 119 and *passim*).

In welcome contrast with familiar Austrian maundering about methodology, Rothbard makes sensible remarks, smoothly working them into his appraisals of earlier writers. Nassau Senior gets good marks. Bad ones go to Sir William Petty, the supposed father of econometrics, and to David Ricardo and John Stuart Mill. Readers should not be misled by Ludwig von Mises's unusual use of the words *a priori* in characterizing economic theory. Economics deduces its theorems from axioms, axioms of such pervasive validity and crushing obviousness that we can scarcely imagine persons, societies, and worlds of which they were not true. (For example, resources are scarce relative to almost unlimited wants; choices must be made; and people act purposefully, taking account of the expected consequences of

alternative courses of action. Besides scattered passages in these two volumes, see Rothbard's article in the *Southern Economic Journal*, January 1957).

If I had an opportunity to do so—which, sadly, none of us ever will—I would quibble with Rothbard on several points. He describes the function of the landlord as that of deciding or administering the allocation of pieces of land among rival uses (II, p. 91). This view seems too narrow to explain the huge rents that some landlords collect; it requires integration with capital and interest theory. As a capitalist, the landlord collects something closely akin to interest for waiting, meaning the tying up over time of wealth that might otherwise have been used to bid resources into serving his consumption or other near-term purposes. At least potentially, furthermore, the landlord organizes the provision of spatially confined civic goods, often called public goods (Fred Foldvary, *Public Goods and Private Communities* [Aldershot, England and Brookfield, Vt.: Edward Elgar, 1994]). I'd like to discuss Rothbard's evident complacency about monetary deflation (e.g., II, p. 174) and his attacks on utilitarianism (esp. II, chap. 2).

I am approaching a conclusion. Rothbard's riding his hobbies contributes to the charm of his work. Although not evidently religious himself, he took interest in religion as a social phenomenon and as an influence on thinking about secular matters. He may occasionally have been carried away, as in devoting a dozen or so pages to the Anabaptists of the sixteenth century. Yet even this digression is not entirely irrelevant to the emergence of economic ideas—the Anabaptists were early communists—and it and similar digressions do help make his history eminently readable.

No one surveying so many thinkers in such detail as Rothbard does could have read every word of their every major work. Reliance on secondary sources was inevitable. Rothbard is frank and generous in recognizing the scholarship of their authors, including ones who did not share his political views. He often cites Joseph Schumpeter, Emil Kauder, Alexander Gray, Marjorie Grice-Hutchinson, John T. Noonan, and F. A. Hayek (whom he treats respectfully, although not without plausible criticism on specific points, as at I, p. 527). I was delighted to see recognition given to his and my Columbia professor Joseph Dorfman and to my Oberlin professor Raymond de Roover. In using secondary sources, Rothbard does not merely pluck from them; he assesses them, as is important for identifying

understandings and misunderstandings about the history of the discipline.

The degree of documentation given for specific statements varies among sections of the two volumes. Even an occasional direct quotation goes without citation. Trivial lapses like these, coupled with my own merely amateur knowledge of Rothbard's subject matter, keep me from authoritatively vouching for the accuracy of his every statement.

Rothbard's work helps underline why economists should study the history of thought. The Whig theory of history as applied to science is wrong—the standard “complacent and infuriating Panglossian optimism” that forgets the real possibility of deterioration. Rothbard shares Thomas Kuhn's “less than starry-eyed view of science,” yet without adopting Kuhn's “nihilistic philosophic outlook.” Especially in economics, “[t]here can . . . be no presumption whatever . . . that later thought is better than earlier” (I, Introduction).

Although not using the term “frontiersmanship,” Rothbard is aware of the phenomenon—the self-congratulatory attitude of workers on the supposed frontiers of economic science who scorn attention to work of earlier generations as mere antiquarianism on the ground that anything worth knowing is already incorporated into their discipline's current wisdom. Unlike the frontiersmen, Rothbard knows that earlier progress can get forgotten. So can earlier fallacies, which keep getting independently rediscovered and commanding attention. Examples are versions of the real-bills doctrine blaming balance-of-payments deficits for the depreciation on the foreign-exchange market of currencies undergoing inflation.

Less so in economics than in the natural sciences do valid discoveries get embodied not only into pure knowledge but also into technology, many of whose users have a profit-and-loss incentive to get things straight. That incentive is notoriously perverted or weak for politicians and special-interest lobbyists. In economics, therefore, we need scholars who specialize in keeping us alert to old contributions—and old fallacies—masquerading as new truths. Specialists in economic thought should also be alert to the danger of an approach that stresses the work of a few great men and strives mightily to interpret their writings as justifying admiration traditionally accorded them. The resulting intellectual atmosphere may well contribute to robbing Rothbard's lesser figures (like Kendall of Kentucky) of the attention that their not-yet-fashionable ideas merited.

Rothbard occasionally (e.g., II, p. 456) mentions economists to be covered or points to be developed in a “later volume” of his mammoth project. These promises have now become poignant. Anyway, Rothbard’s work will suggest research topics and provide inspiration for rising generations of Austrian economists.

Leland B. Yeager
Auburn University

THE **REVIEW OF AUSTRIAN
ECONOMICS**

Volume 9, Number 2

The Review of Austrian Economics

Founding Editor: Murray N. Rothbard (1926–1995)

Editors

Walter Block
College of the Holy Cross
Hans-Hermann Hoppe
University of Nevada, Las Vegas
Joseph T. Salerno
Pace University

Book Review Editor

David Gordon
The Mises Review

Associate Editors

Steve H. Hanke
The Johns Hopkins University
Jeffrey M. Herbener
Washington and Jefferson College

Managing Editor

Judith F. Thommesen

Editorial Assistant

John S. Thompson

Editorial Board

- | | | |
|--|---|--|
| Martin Anderson
<i>Hoover Institution
Stanford University</i> | Lowell Gallaway
<i>Ohio University</i> | William H. Peterson
<i>Washington, D.C.</i> |
| Dominick T. Armentano
<i>East Vero Beach, Florida</i> | Fred R. Glahe
<i>University of Colorado</i> | W. Duncan Reekie
<i>University of the
Witwatersrand
South Africa</i> |
| Roger Arnold
<i>California State
University, San Marcos</i> | Robert F. Hébert
<i>Auburn University</i> | Morgan O. Reynolds
<i>Texas A&M
University</i> |
| Charles W. Baird
<i>California State
University, Hayward</i> | Robert Higgs
<i>Independent Institute</i> | Charles K. Rowley
<i>George Mason
University</i> |
| Don Bellante
<i>University of South
Florida</i> | Ole-Jacob Hoff
<i>Tjøme, Norway</i> | Arthur Seldon
<i>Institute of Economic
Affairs, London</i> |
| James T. Bennett
<i>George Mason
University</i> | Randall G. Holcombe
<i>Florida State University</i> | Sudha Shenoy
<i>University of Newcastle
Australia</i> |
| Olafur Björnsson
<i>University of Iceland</i> | Edward Kaplan
<i>Western Washington
University</i> | Mark Skousen
<i>Rollins College</i> |
| Samuel Bostaph
<i>University of Dallas</i> | Lord Harris of High Cross
<i>Institute of Economic
Affairs, London</i> | Gene Smiley
<i>Marquette University</i> |
| J. B. Bracewell-Milnes
<i>Surrey, England</i> | Kurt Leube
<i>Hoover Institution
Stanford University</i> | John W. Sommer
<i>Urban Institute
University of North
Carolina, Charlotte</i> |
| Yale Brozen
<i>University of Chicago</i> | Stephen O. Littlechild
<i>University of Birmingham
England</i> | T. Alexander Smith
<i>University of Tennessee</i> |
| Alfred G. Cuzan
<i>University of West
Florida</i> | Naomi Moldofsky
<i>University of Melbourne
Australia</i> | Thomas C. Taylor
<i>Wake Forest University</i> |
| Thomas J. DiLorenzo
<i>Loyola College</i> | Thomas Gale Moore
<i>Hoover Institution
Stanford University</i> | Richard K. Vedder
<i>Ohio University</i> |
| Richard M. Ebeling
<i>Hillsdale College</i> | John C. Moorhouse
<i>Wake Forest University</i> | Leland B. Yeager
<i>Auburn University</i> |
| Robert B. Ekelund, Jr.
<i>Auburn University</i> | David O'Mahoney
<i>University College
Cork, Ireland</i> | Albert H. Zlabinger
<i>Carl Menger Institute
Vienna</i> |
| Joseph Fuhrig
<i>Golden Gate
University</i> | E. C. Pasour, Jr.
<i>North Carolina State
University</i> | |

THE REVIEW OF AUSTRIAN ECONOMICS

Volume 9, Number 2

Walter Block, Co-Editor
College of the Holy Cross

Hans-Hermann Hoppe, Co-Editor
University of Nevada, Las Vegas

Joseph T. Salerno, Co-Editor
Pace University

David Gordon, Book Review Editor
The Mises Review

LUDWIG
VON
MISES
INSTITUTE

Auburn University



Kluwer Academic Publishers
Boston/Dordrecht/London

The Review of Austrian Economics

SUBSCRIPTION RATES

The subscription price for Volume 10 (2 issues) is:

For institutions Dfl. 315.00/US \$202.00 including postage and handling.

ORDERING INFORMATION / SAMPLE COPIES

Subscription orders and requests for sample copies should be sent to:

Kluwer Academic Publishers Group
P.O. Box 322
3300 AH Dordrecht
The Netherlands

or

Kluwer Academic Publishers
101 Philip Drive
Assinippi Park
Norwell, MA 02061, USA

or to any subscription agent

All requests for further information concerning *The Review of Austrian Economics* should be sent to:

Kluwer Academic Publishers
101 Philip Drive
Assinippi Park
Norwell, MA 02061, USA

or

Kluwer Academic Publishers
P.O. Box 163
3300 AD Dordrecht
The Netherlands

Photocopying. *In the U.S.A.:* This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Kluwer Academic Publishers for users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the base fee of \$8.50 per copy per article is paid directly to CCC. For those organizations that have been granted a photocopying license by CCC, a separate system of payment has been arranged. The fee code for users of the Transactional Reporting Service is 0889-3047/96/\$8.50.

Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale.

In the rest of the world: Permission to photocopy must be obtained from the copyright owner.

Copyright © 1996 by The Ludwig von Mises Institute.

ISSN 0889-3047

Printed on acid-free paper.

The Review of Austrian Economics has two broadly conceived objectives: (1) to promote the development and extension of Austrian economics and (2) to promote the analysis of contemporary issues in the mainstream of economics from an Austrian perspective.

Correspondence relating to submissions, articles, book reviews, permission to quote, and back issues through volume 3 should be addressed to: *Review of Austrian Economics*, Ludwig von Mises Institute, Auburn University, Auburn, Alabama 36849.

Papers submitted for consideration should be double-spaced, include three copies of the manuscript, and an abstract of not more than 250 words to assist the referees. There is no submission fee. Articles that are accepted will be requested on an IBM format diskette, when possible.

The opinions expressed in this volume by the authors in their respective articles are their own and *The Review of Austrian Economics* assumes no responsibility for these views.

This journal is abstracted and/or indexed in *Sociological Abstracts* and *Social Planning/Policy & Development Abstracts* and IBZ.

THE | REVIEW OF AUSTRIAN ECONOMICS

**Dedicated to the Memory
of Murray N. Rothbard**

From the Editors	1
Peter G. Klein Economic Calculation and the Limits of Organization	3
Pascal Salin Cartels as Efficient Productive Structures	29
Thomas J. DiLorenzo The Myth of Natural Monopoly	43
Jesús Huerta de Soto New Light on the Prehistory of the Theory of Banking and the School of Salamanca	59
George Selgin and Lawrence H. White In Defense of Fiduciary Media—or, We Are <i>Not</i> Devo(lutionists), We Are Misesians!	83
Roger W. Garrison Central Banking, Free Banking, and Financial Crises	109
Richard E. Wagner Who Owes What, and To Whom? Public Debt, Ricardian Equivalence, and Governmental Form	129
Israel M. Kirzner Reflections on the Misesian Legacy in Economics	143



Murray N. Rothbard
1926-1995

From the Editors

This issue is dedicated to the memory of Murray N. Rothbard. Not only was he the founding editor of this journal, but, in recent years, he was also widely and justifiably hailed as the “Dean of the Modern Austrian School.” Indeed, it was his enormous scholarly productivity, whose depth and breadth are clearly unparalleled among contemporary economists, that served as the main inspiration for the extraordinary and ongoing renaissance of interest in Austrian economics that has taken place in the last two decades. As further testimony to the breathtaking range of his scholarly contributions, which extended beyond economics to the social sciences in general, an issue of the interdisciplinary *Journal of Libertarian Studies* honoring his memory is being published contemporaneously.

The articles in this issue of the *Review of Austrian Economics*, contributed by friends and colleagues of Professor Rothbard’s, provide just a small sample of the diverse fields to which he made original contributions. The topics of these articles include the theory of monopoly, the economic function and social-welfare implications of (noncompulsory) cartels, economic calculation as a constraint on the size of the firm, the origin and consequences of fractional-reserve banking, the link between central banking and financial crises, the nature of public debt, the money and banking theory of the sixteenth-century Spanish School of Salamanca, and the precise nature of the Misesian contribution to economics. As editors, we have chosen to publish these articles in the belief that, as scientific contributions in their own right, they will supply continued momentum to the Austrian revival. We can conceive of no greater tribute to the memory of Murray Rothbard.

Walter Block
Hans-Hermann Hoppe
Joseph T. Salerno

The Ludwig von Mises Institute gratefully acknowledges the generosity of its Members, who made the publication of this journal possible. In particular, it wishes to thank the following Patrons:

Mark M. Adamo	Hugh E. Ledbetter
Morgan Adams, Jr.	Winston Lehman
The Adams Fund	Arthur L. Loeb
O. P. Alford, III [†]	Jack E. Magoulakis
Anonymous (3)	David Manley
George E. Atkinson, Jr.	Dr. Walter Marcyan
Donald Bently	Arthur Marron
John Hamilton Bolstad	Forrest E. Mars, Sr.
Philip G. Brumder	William W. Massey, Jr.
Robert H. Clarke	Roy G. Michell, Jr.
Douglas J. Cook	Roy G. Michell Foundation
Cook Brothers Manufacturing	Haydock H. Miller, Jr.
Mr. and Mrs. William C. Daywitt	Frank and Floralie Millsaps
Maino and Dorothy des Granges	M. K. Morse
Richard A. Elwyn	The Nehemiah Foundation
Dr. Larry J. Eshelman	Mr. and Mrs. Brantley I. Newsom
James and Vickie Feutz	Ellis L. Phillips, Jr.
Willard Fischer	Clifton B. Rodes
James W. Frevert	James M. Rodney
John F. Gemmill	Catherine Dixon Roland
Henry Getz	Sheldon Rose
Hager Lumber Company	E. D. Shaw, Jr. [†]
Frank W. Heemstra	Quinten E. Ward
In memory of	David and Barbara Westrate
John A. and Erlene Hendrickson	Jerald N. Wigutow
Albert L. Hillman, Jr.	Thomas G. Winter
John Hotonian	John R. Wright
Hugh P. King	Dr. Carl E. Youngstrom
Robert H. Kriebel	

Economic Calculation and the Limits of Organization

Peter G. Klein

Economists have become increasingly frustrated with the textbook model of the firm. The “firm” of intermediate microeconomics is a production function, a mysterious “black box” whose insides are off-limits to respectable economic theory (relegated instead to the lesser disciplines of management, organization theory, industrial psychology, and the like). Though useful in certain contexts, the textbook model has proven unable to account for a variety of real-world business practices: vertical and lateral integration, geographic and product-line diversification, franchising, long-term commercial contracting, transfer pricing, research joint ventures, and many others. As an alternative to viewing the firm as a production function, economists are turning to a new body of literature that views the firm as an *organization*, itself worthy of economic analysis. This emerging literature is the best-developed part of what has come to be called the “new institutional economics.”¹ The new perspective has deeply enhanced and enriched our understanding of firms and other organizations, such that we can no longer agree with Ronald Coase’s 1988 statement that “[w]hy firms exist, what determines the number of firms, what determines what firms do . . . are not questions of interest to most economists” (Coase 1988a, p. 5). The new theory is not without its critics; Richard Nelson (1991), for example, objects that the new institutional economics tends to downplay discretionary differences among firms. Still, the new institutional economics—in particular, agency theory and transaction cost economics—has been

*Peter G. Klein is assistant professor of economics at the University of Georgia. He thanks (without implicating) Don Boudreaux, Jerry Ellig, Sandy Klein, Dick Langlois, Joe Salerno, and Oliver Williamson for helpful comments and suggestions. Parts of this paper were presented as “Socialism and the Theory of the Firm” at the Institutional Analysis Workshop, University of California, Berkeley.

¹For overviews of the new institutional economics and the theory of the firm, see Coase (1991); Furubotn and Richter (1991); Holmström and Tirole (1989), Langlois (1994b), and *Journal of Institutional and Theoretical Economics* (March 1993, March 1994, and March 1995). For a survey of related empirical work see Shelanski and Klein (1995).

the subject of increasing attention in industrial organization, corporate finance, strategic management, and business history.²

This paper highlights some distinctive Austrian contributions to the theory of the firm, contributions that have been largely neglected, both inside and outside the Austrian literature. In particular, I argue that Mises's concept of economic calculation—the means by which entrepreneurs adjust the structure of production to accord with consumer wants—belongs at the forefront of Austrian research into the nature and design of organizations. There is a unique Austrian perspective on economic planning, a perspective developed over the course of the socialist calculation debate. As was recognized in the early Austrian reinterpretations of the calculation debate (Lavoie 1985, Kirzner 1988), Mises's conception of the problem faced by socialist planners is part and parcel of his understanding of how resources are allocated in a market system. Mises himself emphasized that planning is ubiquitous: “[E]very human action means planning. What those calling themselves planners advocate is not the substitution of planned action for letting things go. It is the substitution of the planner's own plan for the plans of his fellow men” (Mises 1947, p. 493). All organizations plan, and all organizations, public and private, perform economic calculation. In this sense, the calculation problem is much more general than has usually been realized.

With their unique perspective on markets and the difficulties of resource allocation under central planning, third- and fourth-generation Austrian economists have always implicitly understood the economics of organization. In this context, as Nicolai Juul Foss (1994, p. 32) notes in a recent issue of this *Review*, “it is something of a doctrinal puzzle that the Austrians have never formulated a theory of the firm.” Foss points out that many elements of the modern theory of the firm—property rights, relationship-specific assets, asymmetric information, the principal-agent problem—appeared, at least in elementary form, in Austrian writings since the middle stages of the calculation debate. Indeed, Rothbard's treatment of firm size in *Man, Economy, and State* (1962) was among the first discussions to adopt explicitly the framework proposed by Ronald Coase in 1937, a framework that underlies most contemporary theorizing about the firm. Mises's discussion in *Human Action* (1949) of the role of financial markets foreshadows Henry Manne's seminal 1965 article on the market for corporate control along with the recent recognition of finance as an essential part of economics.

²The framework of transaction cost economics has already made it into textbook form: Kreps (1990, pp. 744–90), Rubin (1990), Milgrom and Roberts (1992), Baye and Beil (1994), and Acs and Gerlowski (1996).

Besides anticipating parts of the modern literature, Mises and Rothbard also introduced significant innovations, though this has not yet been generally recognized. Their contributions, while not part of a fully articulated, explicit theory of the firm, deserve attention and development, especially by those working on such issues from within the Austrian School.³ These contributions are Rothbard's application of the calculation problem to the limits of the firm, and Mises's discussion of how the financial markets both limit managerial discretion and perform the ultimate resource allocation task in a market economy.

The Textbook Theory of the Firm

In neoclassical economic theory, the firm as such does not exist at all. The "firm" is a production function or production possibilities set, a means of transforming inputs into outputs. Given the available technology, a vector of input prices, and a demand schedule, the firm maximizes money profits subject to the constraint that its production plans must be technologically feasible. That is all there is to it. The firm is modeled as a single actor, facing a series of relatively uncomplicated decisions: what level of output to produce, how much of each factor to hire, and so on. These "decisions," of course, are not really decisions at all; they are trivial mathematical calculations, implicit in the underlying data. In the long run, the firm may also choose an optimal size and output mix, but even these are determined by the characteristics of the production function (economies of scale, scope, and sequence). In short: the firm is a set of cost curves, and the "theory of the firm" is a calculus problem.

To be sure, these models are not advertised as realistic descriptions of actual business firms; their use is purely instrumental. As David Kreps (1990, p. 233)—himself much less sanguine about the merits of the traditional model than most—puts it: if real-world firms do not maximize profits as the traditional theory holds, "that doesn't mean that profit maximization isn't a good positive model. Only the data can speak to that, and then only after we see the implications of profit maximization for observable behavior." However, even granting instrumentalism its somewhat dubious merits,⁴ the production-function approach is unsatisfactory, because it isn't useful for understanding a variety of economic phenomena. The black-box model is really a theory about a *plant* or production process, not about a *firm*. A single firm can own and operate multiple

³I have in mind recent work by Boudreaux and Holcombe (1989), Foss (1993a, 1993b, 1993c), Langlois (1988, 1992a, 1992b), and Minkler (1993a, 1993b).

⁴For critiques of instrumentalism see Rizzo (1985) and Batemarco (1985). For references to the interpretative literature on Milton Friedman's 1953 essay on "positive economics"—the source of most economists' views on method—see Boland (1979), Caldwell (1980), and Musgrave (1981); all reprinted in Caldwell (1984) along with De Marchi (1988).

production processes. Similarly, two or more firms can contract to operate jointly a single production process (as in a research joint venture). If we want to understand the scale and scope of the firm as a legal entity, then, we must look beyond the textbook model.

Coase and Transaction Costs

Ronald Coase, in his celebrated 1937 paper on “The Nature of the Firm,” was the first to explain that the boundaries of the organization depend not only on the productive technology, but on the costs of transacting business. In the Coasian framework, as developed and expanded by Williamson (1975, 1985, 1996), Klein, Crawford, and Alchian (1978), and Grossman and Hart (1986), the decision to organize transactions within the firm as opposed to on the open market—the “make or buy decision”—depends on the relative costs of internal versus external exchange. The market mechanism entails certain costs: discovering the relevant prices, negotiating and enforcing contracts, and so on. Within the firm, the entrepreneur may be able to reduce these “transaction costs” by coordinating these activities himself. However, internal organization brings another kind of transaction cost, namely problems of information flows, incentives, monitoring, and performance evaluation. The boundary of the firm, then, is determined by the tradeoff, at the margin, between the relative transaction costs of external and internal exchange. In this sense, firm boundaries depend not only on technology, but on organizational considerations; that is, on the costs and benefits of contracting.

The relative costs of external and internal exchange depend on particular characteristics of transaction: the degree to which relationship-specific assets are involved, the amount of uncertainty about the future and about trading partners’ actions, the complexity of the trading arrangement, and the frequency with which the transaction occurs. Each matters in determining the preferred institutional arrangement (that is, internal versus external production), although the first—“asset specificity”—is held to be particularly important.⁵ Williamson (1985, p. 55) defines asset specificity as “durable investments that are undertaken in support of particular transactions, the opportunity cost of which investments are much lower in best alternative uses or by alternative users should the original transaction be prematurely terminated.” This could describe a variety of relationship-specific investments, including both specialized physical and human capital, along with intangibles such as R&D and firm-specific knowledge or capabilities.

⁵Indeed, Williamson’s transaction cost economics is sometimes described as the “governance” or asset-specificity branch of the New Institutional Economics, as opposed to the “measurement” or team production branch (associated with Alchian and Demsetz 1972).

The recent transformation of economists' thinking about the firm has been nicely summarized by Mark Roe (1994, p. vii):

Economic theory once treated the firm as a collection of machinery, technology, inventory, workers, and capital. Dump these inputs into a black box, stir them up, and one got outputs of products and profits. Today, theory sees the firm as more, as a management structure. The firm succeeds if managers can successfully coordinate the firm's activities; it fails if managers cannot effectively coordinate and match people and inputs to current technologies and markets. At the very top of the firm are the relationships among the firm's shareholders, its directors, and its senior managers. If those relationships are dysfunctional, the firm is more likely to stumble.⁶

With this new orientation, economic theory is playing an increasingly visible role in finance, accounting, management, and other areas once thought to be beyond the purview of economics.

Economic Calculation and the Limits to Firm Size

Unfortunately, the growing economics literature on the theory of the firm focuses mostly on the costs of market exchange, and much less on the costs of governing internal exchange. The new research has yet to produce a fully satisfactory explanation of the limits to firm size (Williamson 1985, chap. 6). In Coase's words, "Why does the entrepreneur not organize one less transaction or one more?" Or, more generally, "Why is not all production carried on in one big firm?" (Coase 1937, pp. 42–43). The theory of the limits to the firm is perhaps the most difficult and least well developed part of the new economics of organization. Existing contractual explanations rely on problems of authority and responsibility (Arrow 1974); incentive distortions caused by residual ownership rights (Grossman and Hart 1986; Holmström and Tirole 1989); and the costs of attempting to reproduce market governance features within the firm (Williamson 1985, chap. 6). It is here that Austrian theory has an obvious contribution to make, by applying Mises's theorem on the impossibility of economic calculation under socialism. Rothbard has shown how the need for monetary calculation in terms of actual prices not only explains the failures of central planning under socialism, but places an upper bound on firm size.

The Socialist Calculation Debate: A Brief Review

To understand Mises's position in the calculation debate, one must realize that his argument is not exclusively, or even primarily, about

⁶Austrians would add that capital, land, and labor—"management" included—are not the only inputs or factors of production. There is also *entrepreneurship* or uncertainty bearing, and what Rothbard (1962, pp. 538–41) calls *ownership* or the "decision-making factor." On this see also Mises (1949, pp. 291–52, and pp. 66–68) below.

socialism. It is about the role of prices for capital goods. Entrepreneurs make decisions about resource allocation based on their expectations about future prices, and the information contained in present prices. To make profits, they need information about all prices, not only the prices of consumer goods but the prices of factors of production. Without markets for capital goods, these goods can have no prices, and hence entrepreneurs cannot make judgments about the relative scarcities of these factors. In short, resources cannot be allocated efficiently. In any environment, then—socialist or not—where a factor of production has no market price, a potential user of that factor will be unable to make rational decisions about its use. Stated this way, Mises's claim is simply that efficient resource allocation in a market economy requires well-functioning asset markets. Because scholars differ about what Mises "really meant," however, it may be useful here to provide a brief review of the debate.

Before 1920, according to the standard account,⁷ socialist theorists paid little attention to how a socialist economy would work in practice, most heeding Marx's admonition to avoid such "utopian" speculation. Then Mises, known at the time mainly as a monetary theorist, published the sensational article later translated as "Economic Calculation in the Socialist Commonwealth" (1920).⁸ Mises claimed that without private ownership of the means of production, there would be no market prices for capital goods, and therefore no way for decisionmakers to evaluate the relative efficiency of various production techniques. Anticipating the later argument for "market socialism," Mises argued that even if there were markets for consumer goods, a central planner could not "impute" meaningful prices to capital goods used to produce them. In short, without market-generated prices for both capital and consumer goods, even the most dedicated planner would find it "impossible" to allocate resources according to consumer wants.

Throughout the 1920s and early 1930s Mises's argument became the focus of intense discussion within the German-language literature. Eventually it was agreed that Mises was correct at least to point out that a socialist society could not do without such things as money and prices, as some early socialists had suggested, and that there was no feasible way to set prices according, say, to quantities of labor time. Nonetheless, it was felt that Vilfredo Pareto and his follower Enrico Barone (1908) had shown that nothing was "theoretically" wrong with

⁷For examples of the "standard account" of the calculation debate see Schumpeter (1942, pp. 172–86) and Bergson (1948). My discussion of the "revisionist view" follows Hoff (1949), Salerno (1990), and Rothbard (1991).

⁸Other works that made arguments similar to that of Mises include N. G. Pierson's "The Problem of Value in the Socialist Community" (1902) and parts of Max Weber's *Economy and Society* (1921).

socialism, because the requisite number of demand and supply equations to make the system “determinate” would exist under either capitalism or socialism. If the planners could somehow get the necessary information on preferences and technology, they could in principle compute an equilibrium allocation of final goods.

The most important response to Mises, however, and the one almost universally accepted by economists, was what became known as “market socialism” or the “mathematical solution,” developed by Fred Taylor (1929), H. D. Dickinson (1933), Abba Lerner (1934), and Oskar Lange (1936–37). In a system of market socialism, capital goods are collective property, but individuals are free to own and exchange final goods and services. The system would work like this. First, the Central Planning Board chooses arbitrary prices for consumer and capital goods. At those prices, the managers of the various state-owned enterprises are instructed to produce up to the point where the marginal cost of each final good is equal to its price, and then to choose the input mix that minimizes the average cost of producing that quantity. Then, consumer goods prices are allowed to fluctuate, and the Central Planning Board adjusts the prices of capital goods as shortages and surpluses of the final goods develop. Resources would thus be allocated according to supply and demand, through a process of “trial-and-error” essentially the same as that practiced by the managers of capitalist firms. Lange’s contribution, it has generally been held, was to show that production under market socialism could be just as efficient as production under capitalism, since the socialist planners “would receive exactly the same information from a socialized economic system as did entrepreneurs under a market system” (Heilbroner 1970, p. 88).⁹

Market socialism was seen as an answer not only to Mises’s calculation problem, but also to the issue of “practicality” raised by Hayek and Lionel Robbins. Hayek, in his contributions to *Collectivist Economic Planning* (Hayek, ed., 1935a), later expanded in “The Competitive Solution” (1940) and his well-known papers “Economics and Knowledge” (1937) and “The Use of Knowledge in Society” (1945), and Robbins, in his *The Great Depression* (1934), had changed the terms of the debate by focusing not on the problem of calculation, but on the problem of knowledge. For Hayek and Robbins, the failure of socialist organization is due to a mechanism design problem, in that planners cannot allocate resources efficiently because they cannot obtain complete information on consumer preferences and resource availability. Furthermore, even if the planners were

⁹It would no doubt be gratuitous to point out that since the collapse of central planning in Eastern Europe the writer of that comment has changed his mind, writing that although “[f]ifty years ago, it was felt that Lange had decisively won the argument for socialist planning,” now “[i]t turns out, of course, that Mises was right” (Heilbroner 1990, p. 92).

somehow able to acquire these data, it would take years to compute the millions of prices used by a modern economy. The Lange–Lerner–Taylor approach claimed to solve this preference-revelation problem by trial-and-error, so no actual computations would be necessary.¹⁰

With the widespread acceptance of the theory of market socialism, there developed an “orthodox line” on the socialist calculation debate, neatly summarized in Abram Bergson’s well-known survey of “Socialist Economics” (1948) and in Joseph Schumpeter’s *Capitalism, Socialism and Democracy* (1942, pp. 172–86). According to this line, Mises first raised the problem of the possibility of economic calculation under socialism, only to be refuted by Pareto and Barone; Hayek and Robbins then “retreated” to the position that socialist planners could calculate in theory, but that in practice the information problem would make this too difficult; then the market socialists showed that trial and error would eliminate the need for complete information on the part of the planners. Therefore, the argument goes, economic theory *per se* can say nothing conclusive about the viability of central planning, and the choice between capitalism and socialism must be purely political.

Calculation versus Incentives

The orthodox line on socialist planning has been modified in recent years with the development of incentive and information theory. The differences between capitalism and socialism, it is now typically held, lie in the different *incentive* properties of the two systems. Centrally directed systems are thought to be subject to greater agency costs—managerial discretion, shirking, and so on—than market systems (see, for example, Winiecki 1990). After all, Lange himself warned that “*the real danger of socialism is that of a bureaucratization of economic life*” (Lange 1936–37, p. 109; italics in original).

As has been pointed out elsewhere (Rothbard 1991, pp. 51–52), however, the calculation debate was *not* primarily about agency or managerial incentives. The incentive problem had long been known¹¹

¹⁰Lange actually claimed years later that even market socialism would be made obsolete with the advent of high-speed computers, which could instantly solve the huge system of simultaneous equations for the central planner. “Were I to rewrite my [1936] essay today my task would be much simpler. My answer to Hayek and Robbins would be: So what’s the trouble? Let us put the simultaneous equations on an electronic computer and we shall obtain the solution in less than a second. The market process with its cumbersome *tâtonnements* appears old fashioned. Indeed, it may be considered as a computing device of the pre-electronic age” (Lange 1965, pp. 401–2). Obviously, Lange did not have much experience with a computer. Also, during his time as chairman of the Polish Economic Council in the 1950s, Lange never tried to put market socialism into practice (see Lange 1958).

¹¹We tend to forget just how old the idea of socialism is, that it is not a twentieth-century invention; the subtitle of Alexander Gray’s famous book *The Socialist Tradition* (1946) is “Moses to Lenin.”

(if not fully developed) and was expressed in the famous question: "Under socialism, who will take out the garbage?" That is, if everyone is compensated "according to his needs," what will be the incentive to do the dirty and unpleasant tasks; or, for that matter, any tasks at all? The traditional socialist answer was that self-interest is a *product* of capitalism, and that socialism would bring about a change in human nature. In the worker's paradise would emerge a "New Socialist Man," eager to serve and motivated only by the needs of his fellows. These early theorists seem to have assumed, to borrow the expression used by Oliver Williamson (1991, p. 18) in a critique of a more recent socialist proposal, "the abolition of opportunism by agencies of the state." Experience has exposed the charming naiveté of such notions.

But Mises's challenge to socialism is distinct from this well-known incentive problem.¹² Assume for the moment that everyone is willing to work just as hard under central direction as under a market system. There still remains the problem of exactly what directives the Central Planning Board will issue. The Board will have to decide what goods and services should be produced, how much of each to produce, what intermediate goods are needed to produce each final good, and so on. In a complex, modern economy with multiple stages of production, resource allocation requires the existence of money prices for capital goods, prices that under capitalism arise from an ongoing process of competitive bidding by entrepreneurs for the factors of production. This process cannot be replicated by input-output analysis, computer simulations, or any other form of artificial market. Mises's main point was that socialism fails because decision makers require meaningful prices for all of these factors to choose from the vast array of possible factor combinations. "Without recourse to calculating and comparing the benefits and costs of production using the structure of monetary prices determined at each moment on the market, the human mind is only capable of surveying, evaluating, and directing production processes whose scope is drastically reduced to the compass of the primitive household economy" (Salerno 1990, p. 52).

The distinction between calculation and incentives is important because the modern economics literature on organizational design—from transaction cost explanations of firm size, to public choice theories of bureaucracy, to recent work on market socialism and the "soft budget constraint" (Kornai 1986)—focuses primarily on incentive problems (possibly encouraged by Lange's famous warning about bureaucracy). Incentive theory asks how, within a specified relationship, a principal can get an agent to do what he wants him to do. Mises's problem, however,

¹²Mises does devote a section of the 1920 paper to "Responsibility and Initiative in Communal Concerns," but he clearly considers this a secondary problem for socialist planners, not the primary one.

was different: How does the principal know *what* to tell the agent to do? That is, just what activities ought to be undertaken? What investments should be made? Which product lines expanded and which ones contracted? The ideas developed in the calculation debate suggest that when organizations are large enough to conduct activities that are exclusively internal—so that no reference to the outside market is available—they will face a calculation problem as well as an incentive problem.

In this sense, market-socialist proposals are mostly irrelevant to the real problems of socialist organization. This is the case Mises himself sought to make in his critique of market socialism in *Human Action* (Mises 1949, pp. 694–711). There he complained that the market socialists—and, for that matter, all general equilibrium theorists—misconceive the nature of “the economic problem.” Lange, Lerner, and Taylor looked primarily at the problem of consumer goods pricing, while the crucial problem facing a modern economy concerns the capital structure: namely, in what way should capital be allocated to various activities? The market economy, Mises argued, is driven not by “management”—the performance of specified tasks, within a framework given to the manager—but by *entrepreneurship*, meaning the speculation, arbitrage, and other risk-bearing activities that determine just what the managerial tasks are. It is not managers but entrepreneurs, acting in the capital and money markets, who establish and dissolve corporations, create and destroy product lines, and so on. These are precisely the activities that even market socialism seeks to abolish. In other words, to the extent that incentives are important, what socialism cannot preserve is high-powered incentives not in management, but in entrepreneurial forecasting and decisionmaking.

Mises has been described as saying that it is unreasonable to expect managers of socialist enterprises to “play market,” to act as if they were managers of private firms where their own direct interests were at stake. This may be true, but Mises’s prime concern was that *entrepreneurs* cannot be asked to “play speculation and investment” (Mises 1949, p. 705). The relevant incentive problem, he maintains, is not that of the subordinate manager (the agent), who takes the problem to be solved as given, but that of the speculator and investor (the principal), who decides just what is the problem to be solved. Lange, Lerner, and Taylor see the market through a strictly static, neoclassical lens, where all the parameters of the system are given and only a computational problem needs to be solved. In fact the market economy is a dynamic, creative, evolving *process*, in which entrepreneurs—using economic calculation—make industries grow and shrink, cause new and different production methods to be tried and others withdrawn, and constantly change the range of available products. It is *these* features of market capitalism, and not the incentives of agents to work hard, that are lost without private property ownership.

Indeed, traditional command-style economies, such as that of the former U.S.S.R., appear to be able only to mimic those tasks that market economies have performed before; they are unable to set up and execute original tasks.

The [Soviet] system has been particularly effective when the central priorities involve catching up, for then the problems of knowing what to do, when and how to do it, and whether it was properly done, are solved by reference to a working model, by exploiting what Gerschenkron . . . called the “advantage of backwardness.” . . . Accompanying these advantages are shortcomings, inherent in the nature of the system. When the system pursues a few priority objectives, regardless of sacrifices or losses in lower priority areas, those ultimately responsible cannot know whether the success was worth achieving. The central authorities lack the information and physical capability to monitor all important costs—in particular opportunity costs—yet they are the only ones, given the logic of the system, with a true interest in knowing such costs. (Ericson 1991, p. 21)

Without economic calculation, there is no way to figure out if tasks have been performed efficiently. Hence without markets for physical and financial capital—which determine what tasks will be performed and whether they have been performed adequately—an economic system has difficulty generating anything *new*, and must rely on outside references to tell it what to do. Of course, the only reason the Soviet Union and the communist nations of Eastern Europe could exist at all is that they never fully succeeded in establishing socialism worldwide, so they could use world market prices to establish implicit prices for the goods they bought and sold internally (Rothbard 1991, pp. 73–74). In Mises’s words, these economies

were not isolated social systems. They were operating in an environment in which the price system still worked. They could resort to economic calculation on the ground of the prices established abroad. Without the aid of these prices their actions would have been aimless and planless. Only because they were able to refer to these foreign prices were they able to calculate, to keep books, and to prepare their much talked about plans. (Mises 1949, pp. 698–99)

As we will see below, the firm is in the same situation: it needs outside market prices to plan and evaluate its actions.

Rothbard and the Limits of Organization

Rothbard’s main contribution to the theory of the firm was to generalize Mises’s analysis of the problem of resource allocation under socialism to

the context of vertical integration and the size of the organization. Rothbard writes in *Man, Economy, and State* that up to a point, the size of the firm is determined by costs, as in the textbook model. But “the ultimate limits are set on the relative size of the firm by the necessity for *markets* to exist in every factor, in order to make it possible for the firm to calculate its profits and losses” (Rothbard 1962, p. 536). This argument hinges on the notion of “implicit costs.” The market value of opportunity costs for factor services—what Rothbard calls “estimates of implicit incomes”—can be determined only if there are external markets for those factors (pp. 542–44). For example, if an entrepreneur hires himself to manage his business, the opportunity cost of his labor must be included in the firm’s costs. But without an actual market for the entrepreneur’s managerial services, he will be unable to figure out his opportunity cost; his balance sheets will therefore be less accurate than they would if he could measure his opportunity cost.

The same problem affects a firm owning multiple stages of production. A large, integrated firm is typically organized as groups of semi-autonomous business units or “profit centers,” each unit or division specializing in a particular final or intermediate product. The central management of the firm uses the implicit incomes of the business units, as reflected in statements of divisional profit and loss, to allocate physical and financial capital across the divisions. More profitable divisions are expanded, while less profitable divisions are scaled back. Suppose the firm has an upstream division selling an intermediate component to a downstream division. To compute the divisional profits and losses, the firm needs an economically meaningful “transfer price” for the component. If there is an external market for the component, the firm can use that market price as the transfer price.¹³ Without a market price, however, a transfer price must be estimated in another way.

In practice, this is typically done on a cost-plus basis; sometimes, the buying and selling divisions are left free to bargain over the price (Eccles and White 1988; Shelanski 1993; King 1994). At the very least, any artificial or substitute transfer prices will contain less information than actual market prices; Rothbard (1962, p. 547) puts it more strongly, calling a substitute price “only an arbitrary symbol.” In either case, firms relying

¹³Rothbard (1962, pp. 900–1, n. 56) notes that the implicit transfer price may be somewhat more or less than the existing market price, since the entry of either the buying or the selling division into the external market may bid the price up or down slightly. Unlike Hirshleifer (1956), then, Rothbard does not require the external market to be perfectly competitive for a market-based transfer price to be economically meaningful. For Rothbard, “thin” markets are adequate: all that is necessary to have a genuine “external market” is the existence of at least one other producer (seller) of the intermediate good.

Of course, if external prices are perfectly competitive, then the economy must be in a competitive general equilibrium, in which information is perfect and all contracts are complete, and in which there is thus no need for firms.

on these prices will suffer. "Not being able to calculate a price, the firm could not rationally allocate factors and resources from one stage [or division] to another" (p. 547). The use of internally traded intermediate goods for which no external market reference is available introduces distortions that reduce organizational efficiency. This gives us the element missing from contemporary theories of economic organization, an upper bound: the firm is constrained by the need for external markets for all internally traded goods. In other words, no firm can become so large that it is both the unique producer and user of an intermediate product; for then no market-based transfer prices will be available, and the firm will be unable to calculate divisional profit and loss and therefore unable to allocate resources correctly between divisions. As Rothbard puts it:

Since the free market always tends to establish the most efficient and profitable type of production (whether for type of good, method of production, allocation of factors, or size of firm), we must conclude that complete vertical integration for a capital-good product can never be established on the free market (above the primitive level). *For every capital good, there must be a definite market in which firms buy and sell that good.* It is obvious that this economic law *sets a definite maximum to the relative size of any particular firm on the free market.* . . . Economic calculation becomes ever more important as the market economy develops and progresses, as the stages and the complexities of type and variety of capital goods increase. Ever more important for the maintenance of an advanced economy, then, is the preservation of *markets* for all the capital and other producers' goods. (pp. 547–48; italics in original)

Like the centrally planned economy, the firm needs market signals to guide its actions; without them the firm cannot survive. Note that in general, Rothbard is making a claim only about the upper bound of the firm, not the incremental cost of expanding the firm's activities (as long as external market references are available). As soon as the firm expands to the point where at least one external market has disappeared, however, the calculation problem exists. The difficulties become worse as more and more external markets disappear, as "islands of noncalculable chaos swell to the proportions of masses and continents. As the area of incalculability increases, the degrees of irrationality, misallocation, loss, impoverishment, etc., become greater" (p. 548). In other words, the firm is limited by the extent to which markets exist for the goods it allocates internally. Without market prices for these goods, the firm must rely on relatively costly and inefficient methods of generating its own accounting prices, to perform internal calculations.¹⁴

¹⁴This does not mean that because external prices are necessary for large firms to

Significantly, it is at this point in the discussion in *Man, Economy, and State* (p. 548) that Rothbard launches into a discussion of the socialist calculation debate, making it obvious that the two issues are inextricably linked. The reason that a socialist economy cannot calculate is not that it is socialist, but because a single agent owns and directs all resources. Expanding on this point in his 1976 essay on “Ludwig von Mises and Economic Calculation Under Socialism,” Rothbard explains:

There is one vital but neglected area where the Mises analysis of economic calculation needs to be expanded. For in a profound sense, the theory is not about socialism at all! Instead, it applies to *any* situation where one group has acquired control of the means of production over a large area—or, in a strict sense, throughout the world. On this particular aspect of socialism, it doesn't matter whether this unitary control has come about through the coercive expropriation brought about by socialism or by voluntary processes on the free market. For what the Mises theory focuses on is not simply the numerous inefficiencies of the political as compared to the profit-making market process, but the fact that a market for capital goods has disappeared. This means that, just as Socialist central planning could not calculate economically, no One Big Firm could own or control the entire economy. The Mises analysis applies to any situation where a market for capital goods has disappeared in a complex industrial economy, whether because of socialism or because of a giant merger into One Big Firm or One Big Cartel. (Rothbard 1976, p. 75)

The Mises analysis thus applies to any situation where the market for a particular capital good disappears because a firm has become so large that it is the unique producer and user of that capital good. As we have seen, such a firm will not be viable.

It is surprising that Rothbard's extension of Mises's argument has received virtually no attention in the Austrian literature, even though the point appears four times in *Man, Economy, and State* (p. 536, p. 543, pp. 547–48, and p. 585) and again in the 1976 essay.¹⁵ The argument

function efficiently, firms will necessarily become larger where external markets are “thick” or better developed. On the contrary, large firms typically arise precisely where external markets are poorly developed or hampered by government intervention; these are the kinds of circumstances that give entrepreneurs an advantage in coordinating activities internally. However, such firms are still constrained by the need for *some* external market reference.

¹⁵Lavoie briefly notes the Rothbard analysis in his *Rivalry and Central Planning* (1985, p. 62n). Fritz Machlup, in a comment on Rothbard's 1976 essay, says he is “intrigued” by the analogy between the central planner's problem and the firm's problem, calling it “an issue I have tried to sell in several of my publications . . . but unfortunately not with sufficient success” (Machlup 1976, p. 114). He cites an early book (Machlup 1934, esp. pp. 209–14) and a later article (Machlup 1974, esp. pp. 42–45 and 52–54), both published in German, on the problem of “artificial” transfer prices. The argument is also foreshadowed by Hayek in *Prices and Production* (1935b, p. 63) in a discussion on vertical integration.

needs further development and elaboration, which should prove a useful exercise because the contemporary literature on the size of the firm lacks an adequate explanation for the limits to organization. The Rothbard analysis also suggests a line of research in business strategy: all else equal, firms able to use market-based transfer prices should outperform, in the long run, firms using administered or negotiated transfer prices.¹⁶ As of yet, there is little empirical work on this topic, despite the possible emergence of an “Austrian school of strategy” (Jacobson 1992). A related issue that has received considerable attention, however, is the difficulty of allocating overhead or fixed cost across divisions. If an input is essentially indivisible (or nonexcludable), then there is no way to compute the opportunity cost of just the portion of the input used by a particular division (see Rogerson 1992, for a discussion of these problems).¹⁷ Firms with high overhead costs should thus be at a disadvantage relative to firms able to allocate costs more precisely between business units. Indeed, in the literature on cost accounting there has been some recent interest in “market simulation accounting” (Staubus 1986), by which firms try to assess the price at which an asset would trade in an active market, based on observed market prices and related information. The Rothbardian position on the limits to firm size suggests that the market simulation approach may prove a useful accounting technique.

By the time of the 1976 paper, Rothbard had adopted an explicitly Coasian framework in his discussion of the limits to firm size. His own treatment, Rothbard says,

serves to extend the notable analysis of Professor Coase on the market determinants of the size of the firm, or the relative extent of corporate planning within the firm as against the use of exchange and the price mechanism. Coase pointed out that there are diminishing benefits and increasing costs to each of these two alternatives, resulting, as he put it, in an “‘optimum’ amount of planning” in the free market system. Our thesis adds that the costs of internal corporate planning become prohibitive as soon as markets for capital goods begin to disappear, so that the free-market optimum will always stop well short not only of One Big Firm throughout the world market but also of *any* disappearance of specific markets and hence of economic calculation in that product or resource. (Rothbard 1976, p. 76)

¹⁶This line of reasoning has interesting implications for the study of innovation. Since the innovating firm is more likely to be using unique intermediate goods, innovation carries with its benefits the cost of more severe internal distortions. Economic calculation is then another obstacle the innovator must overcome.

¹⁷Mises (1944, p. 32) recognized the problem of allocating overhead costs, mentioning this as a possible exception to the notion that divisional accounting costs can reflect “true” costs.

This is noteworthy because even as late as 1972, Coase was describing his 1937 paper as “much cited and little used” (Coase 1972, p. 62). Alchian and Demsetz’s “Production, Information Costs, and Economic Organization” came out only in 1972, and Williamson’s *Markets and Hierarchies* in 1975. Rothbard was thus among the earliest writers to develop and extend the Coasian perspective.

Managerial Discretion and the Financial Markets¹⁸

As mentioned above, much current research in the theory of the firm focuses on the agency problem. Under what conditions can managers exercise discretionary behavior? What kinds of rules, or mechanisms, can be designed to align the manager’s interest with the owner’s? Without effective rules, what actions will managers choose? An early application was the alleged “separation of ownership and control” in the modern corporation. Berle and Means (1932) argued that the modern firm is run not by its owners, the shareholders, but by salaried managers, whose interests are different from those of shareholders and include executive perks, prestige, and similar rewards. If the corporation is diffusely held, no individual shareholder has sufficient motivation to engage in (costly) monitoring of managerial decisions, and therefore discretion will flourish at the expense of the market value of the firm.

Henry Manne’s essay, “Mergers and the Market for Corporate Control” (1965), responded that managerial discretion will be limited as long as there is an active market for control of corporations. When managers engage in discretionary behavior, the share price of the firm falls, and this invites takeover and subsequent replacement of incumbent management. Hence while managers may hold considerable autonomy over the day-to-day operations of the firm, the stock market places strict limits on their behavior.¹⁹ To be sure, there is a large and divergent literature on the effectiveness of the takeover mechanism in providing managerial discipline (see Romano 1992 for a summary). If managers desire acquisitions to increase their own prestige or span of control—to engage in “empire building”—then an unregulated market will generate “too many takeovers.” Other critics point out that if the difference between the current (undervalued) price of the firm and its after-takeover market value is common knowledge, then the target firm’s shareholders will refuse to tender their shares until the current

¹⁸This section is based on Klein (1994, pp. 397–98).

¹⁹There are other mechanisms to limit managers’ discretionary activities, such as the market for managers itself; on this see Fama (1980). Williamson (1975) argues that the capital market, as an outside control device, tends to be less effective than an internal device, such as the adoption of the “M-form” structure. Fama’s article, along with Manne’s and several other important papers on this topic, is collected in Putterman, ed. (1986).

price is bid up, appropriating a share of the returns to the acquiring firm's shareholders. Under those conditions, the market will generate "too few" takeovers.²⁰

The central insight of Manne's paper is also found in Mises's *Human Action* (1949), in the passage distinguishing what Mises calls "profit management" from "bureaucratic management" (pp. 308–11). It is true, Mises acknowledges, that the salaried managers of a corporation hold considerable autonomy over the day-to-day operations of the firm. Nonetheless, the shareholders make the ultimate decisions about allocating resources to the firm, in their decisions to buy and sell stock:

[The Berle–Means] doctrine disregards entirely the role that the capital and money market, the stock and bond exchange, which a pertinent idiom simply calls the "market," plays in the direction of corporate business. . . . [T]he changes in the prices of common and preferred stock and of corporate bonds are the means applied by the capitalists for the supreme control of the flow of capital. The price structure as determined by the speculations on the capital and money markets and on the big commodity exchanges not only decides how much capital is available for the conduct of each corporation's business; it creates a state of affairs to which the managers must adjust their operations in detail. (p. 303)

Mises does not identify the takeover mechanism *per se* as a means for capitalists to exercise control—takeovers were much less popular before the late 1950s, when the tender offer began to replace the proxy contest as the acquisition method of choice—but the main point is clear: The true basis of the market system is not the product market, the labor market, or the managerial market, but the capital market, where entrepreneurial judgments are exercised and decisions carried out.²¹

²⁰The flaw in the latter argument is that shareholders will *not* in general have the same information as incumbent managers, outside "raiders," and other specialists. It is not in the small shareholder's interest to learn these details; that is why he delegates such responsibilities to managers in the first place. The raider who discovers a difference between a firm's current market value and its potential value under new management has an opportunity for an entrepreneurial profit (less the transaction costs of takeover). Because shareholders have delegated these responsibilities, they will not in general earn a share of this profit. As Rothbard (1962, p. 372) observes, however, since shareholders (owners) *choose* to delegate operational responsibility to managers—contracting out, if you will, for the managerial function—they themselves retain the ultimate rights of corporate control.

²¹Compare Rothbard (1962, p. 538): "Hired managers may successfully direct production or choose production processes. But the ultimate responsibility and control of production rests inevitably with the *owner*, with the businessman whose property the product is until it is sold. It is the owners who make the decision concerning how much capital to invest and in what particular processes. And particularly, it is the *owners* who must choose the managers. The ultimate decisions concerning the use of their property and the choice of the men to manage it must therefore be made by the owners and by no one else."

As discussed above, Mises's treatment of the importance of financial markets is also the key to his final rebuttal in *Human Action* to Lange, Lerner, and the other market-socialist critics of his calculation argument (Mises 1949, pp. 694–711). The market socialists, he argued, fail to understand that the main task performed by a market system is not the pricing of consumer goods, but the allocation of capital among various branches of industry. By focusing on production and pricing decisions within a *given* structure of capital, the socialists ignore the vital role of capital markets. Rothbard (1993) notes that the same criticism can be applied to the textbook, production-function model of the firm, where capital is also taken for granted. “Neoclassical microtheory talks about ‘managers’ producing up to the point where $MR=MC$, without ever talking about who or what is allocating capital to them. In short, neoclassical firms are implicitly assumed to have a fixed amount of capital allocated to them . . . and they can *only* use that capital to invest in their own firm and nowhere else. Hence, the nonsensical conclusion that each firm’s manager will try to squeeze out the last cent of profit, pushing production until $MR=MC$.” Fortunately, the new literature on transaction-cost determinants of contractual relations has begun to bring capital back into the received microtheory.

Finally, on the subject of the Berle–Means doctrine, Mises notes in *Human Action* that “the emergence of an omnipotent managerial class is not a phenomenon of the unhampered market economy,” but a result of government policy (Mises 1949, p. 307). Here he expands upon his earlier analysis in *Bureaucracy* (1944), where he attacks the claim that bureaucracy follows naturally from firm size. Mises conceives of bureaucracy as rule-following, as opposed to profit-seeking, behavior. He reserves the term “bureaucratic management” for the governing of activities that have no cash value on the market. As long as a firm’s inputs and outputs are bought and sold, the central management of the firm will have the information provided by market prices to evaluate the efficiency of the various branches and divisions within the firm. Then subordinate managers can be given wide discretion to make daily operational decisions.²² If an organization produces a good or service that has no market price—the output of a government agency, for example—then subordinate managers must be given specific instructions for how to perform their tasks.

²²Chapter 1 of *Bureaucracy*, on profit management and the sources of entrepreneurial profit, contains a remarkably lucid account of economic calculation under capitalism and its impossibility under socialism. “To the entrepreneur of capitalist society a factor of production through its price sends out a warning: Don’t touch me, I am earmarked for another, more urgent need. But under socialism these factors of production are mute” (Mises 1944, p. 29).

Mises also provides a very Coase-like discussion of the make-or-buy decision, though without citation (p. 33).

The fact that managers in a private firm have latitude to make day-to-day decisions, Mises argues, does not make the firm “bureaucratic.” “[N]o profit-seeking enterprise, no matter how large, is liable to become bureaucratic provided the hands of its management are not tied by government interference. The trend toward bureaucratic rigidity is not inherent in the evolution of business. It is an outcome of government meddling with business” (Mises 1944, p. 12). By this Mises means that government interference impedes the entrepreneur’s use of economic calculation and the attempt to use prices to impose managerial discipline. Mises gives three examples (pp. 64–73): taxes and price regulations that interfere with corporate profits (distorting an important signal of managerial performance); laws that interfere with hiring and promotion (including the need to hire public relations staffs and legal and accounting personnel to comply with government reporting requirements); and the omnipresent threat of arbitrary antitrust or regulatory activity, in response to which entrepreneurs must become adept at “diplomacy and bribery” (p. 72).

Mark Roe (1994) develops a similar argument in his recent work on the politics of corporate finance. The phenomenon he calls “strong managers, weak owners” is not an outgrowth of the market process; it is the result of legal restrictions on firm ownership and control. In the U.S., for example, banks and other institutions are forbidden from owning firms; antitrust laws prohibit industrial combinations like the Japanese *keiretsu*; and anti-takeover restrictions dilute the disciplinary effects of the takeover mechanism. Laws that require diffuse ownership create what Roe terms the “Berle–Means corporation,” in which “fragmented ownership shifts power in the firm to managers” (p. 93).²³ Absent such legal restrictions, Mises would argue, managerial autonomy is no inefficiency; it’s an essential tool for operating a large, decentralized organization. But the firm must have accurate divisional accounting statements to evaluate managerial performance, and for this it needs the information contained in market prices.

Alternative Austrian Approaches: Knight, Uncertainty, and “Market-Based Management”

Recently, some Austrian economists have suggested that the Coasian framework may be too narrow, too squarely in the general-equilibrium tradition to deal adequately with Austrian concerns (Boudreaux and Holcombe 1989; Langlois 1994a). They contend that the contemporary theory of the firm, following Coase, retains the perspective of static equilibrium analysis and profit maximization over a fixed set of outcomes

²³On the relationship between corporate governance and economic performance, see also Gilson (1995).

with known probabilities. As an alternative, some writers propose the framework in Frank Knight's *Risk, Uncertainty, and Profit* (1921). The Knightian framework, they argue, offers genuine uncertainty, disequilibrium and process analysis, and thus a scope for real entrepreneurship — aspects purportedly more congenial to Austrians. "The Coasian and Knightian theories of the firm deal with the issue [of the existence of firms] from two different vantage points. The Coasian theory takes the inputs and outputs in the firm's production process as given, and models the firm as an organization that acts to minimize the costs of transforming these inputs into outputs. . . . However, in Knight's model, entrepreneurship is the primary role of the firm" (Boudreaux and Holcombe 1989, p. 152). Williamson's transaction cost economics, as characterized by Langlois (1994a, p. 175), does broaden the notion of cost minimization to include transaction costs as well as production costs, but it remains essentially a static exercise with a limited role for expectations: "Seldom does the theory give thought to the possibility that organizational forms may be influenced as much by environments that exist only as future possibilities, imagined or feared."

These descriptions, however, paint with too broad a brush; as Foss (1993c) has recently pointed out, there are "two Coasian traditions." One tradition, the nexus-of-contracts branch associated with Alchian and Demsetz (1972), studies the design of *ex ante* mechanisms to limit shirking when supervision is costly. Here the emphasis is on monitoring and incentives in an (exogenously determined) moral-hazard relationship. The aforementioned criticisms may apply to this branch of the modern literature, but they do not apply to the other tradition, the governance or asset-specificity branch, especially in Williamson's more heterodox formulation. Williamson's transaction cost framework incorporates non-maximizing behavior (bounded rationality); true, "structural" uncertainty or genuine surprise (complete contracts are held not to be feasible, meaning that all *ex post* contingencies cannot be contracted upon *ex ante*); and process or adaptation over time (trading relationships develop over time, typically undergoing a "fundamental transformation" that changes the terms of trade). In short, "at least some modern theories of the firm do not at all presuppose the 'closed' economic universe—with all relevant inputs and outputs being given, human action conceptualized as maximization, etc.—that [some critics] claim are underneath the contemporary theory of the firm" (Foss 1993a, p. 274). Stated differently, one can adopt an essentially Coasian perspective without abandoning the Knightian or Austrian view of the entrepreneur as an uncertainty-bearing, innovating decision maker.²⁴

²⁴Nor do all Coasian perspectives deny the importance of specialized knowledge or routines in determining a firm's capabilities or "core competence." Transaction cost

Similarly, the approach described in this paper differs from that advanced in the recent literature on “market-based management” (Ellig 1993; Ellig and Gable 1993). Market-based management is the philosophy that firm success depends critically on the ability to replicate market-like features within the organization. One of these is “internal markets” for intermediate goods (and services such as financial, legal, accounting, and R&D support) along with the establishment of strict profit-center divisions. Like market prices, these internal prices convey information about local circumstances. Other features include an explicit “mission” or recognition of the firm’s core competence, clearly defined roles and responsibilities for lower-level employees (analogous to property rights in a market economy), employee rewards based on performance (a profit-and-loss system), a well-defined “corporate culture” (customs, behavioral norms), and decentralized decision making.

Underlying market-based management is the team-production or nexus-of-contracts model of the firm advanced by Alchian and Demsetz (1972), supplemented with the “capabilities” theory of Edith Penrose (1959), G. B. Richardson (1972), and David Teece (1980, 1982). This is not the appropriate place for an extended discussion of the capabilities view; suffice it to say that this literature has both its defenders and its detractors.²⁵ The relevant point here is that the literature on market-based management, like other writings in the nexus-of-contracts tradition, mischaracterizes the nature of “planning” within the firm. For example, it attributes to the Coase–Williamson tradition the view that “internal markets are doomed to failure, because the business firm is by nature a command hierarchy” (Ellig 1993, p. 9). The Coasian tradition, however, does not imply that firms do or should adopt a command-and-control structure; on the contrary, as we have already seen, the modern firm will tend to be significantly decentralized, so that managers and workers at all levels of operations can make use of local knowledge. All decisions are not made from above, by executive fiat; the “M-form” corporation described by Williamson and Chandler is a *blend* of market and hierarchy, of centralization and decentralization.

In other words, the entrepreneur does make *some* decisions by “fiat”; the firm is definitely a “taxi,” rather than a “cosmos” (to use Hayek’s esoteric terminology). This does not imply, however, that *all* decisions must be made from the top; we can agree with the market-based management literature that “neither central planning nor command-and-control are

economics, for example, simply holds that the need for *ex post* governance of contracts in the presence of relationship-specific investments, and not “tacit knowledge” *per se*, is the most useful way to think about the boundaries of the firm. For the case that Austrian economics is more compatible with the capabilities literature (for substantive, not only methodological, reasons), see Minkler (1993b) and Langlois (1944a).

²⁵For critical surveys see Conner (1991) and Teece, Pisano, and Shuen (1992).

the defining characteristics of a business firm" (p. 11). Indeed, given competition in the product and factor markets, firms will always tend to select the optimum amount of "market-like" features. The firm's problem, then, is not too much "conscious" planning; the crucial issue is whether these plans are made, and tested, from within a larger market setting. The entrepreneur's plans can be carried out, as we saw above, only when there are definite markets for all internally traded goods or activities. What firms need is not necessarily internal markets, but the information generated by market prices.

Conclusion

The purpose of this paper has been to highlight some Austrian contributions to the theory of the firm and to suggest directions for future research along the same lines. In particular, Rothbard's argument about the need for markets in intermediate goods, and how that places limits on the scale and scope of the organization, deserves further development. This may be a more fruitful exercise than some work in the alternative Austrian traditions.

References

- Acs, Zoltan J., and Daniel A. Gerlowski. 1996. *Managerial Economics and Organization*. Englewood Cliffs, N.J.: Prentice-Hall.
- Alchian, Armen A., and Harold Demsetz. 1972. "Production, Information Costs, and Economic Organization." *American Economic Review* 62: 777–95.
- Arrow, Kenneth J. 1974. *The Limits of Organization*. New York: W. W. Norton.
- Barone, Enrico. 1908. "The Ministry of Production in the Collectivist State." In Hayek, ed., 1935a. Pp. 245–90.
- Batemarco, Robert. 1985. "Positive Economics and Praxeology: The Clash of Prediction and Explanation." *Atlantic Economic Journal* 13 (2): 31–37.
- Baye, Michael, and Richard Beil. 1994. *Managerial Economics and Business Strategy*. Burr Ridge, Ill.: Irwin.
- Bergson, Abram. 1948. "Socialist Economics." In Howard S. Ellis, ed., *A Survey of Contemporary Economics*. Vol. 1. Homewood, Ill.: Richard D. Irwin. Pp. 412–48.
- Berle, Adolph A., and Gardiner C. Means. 1932. *The Modern Corporation and Private Property*. New York: Macmillan.
- Boettke, Peter J., ed. 1994. *The Elgar Companion to Austrian Economics*. Aldershot, U.K.: Edward Elgar.
- Boland, Lawrence A. 1979. "A Critique of Friedman's Critics." *Journal of Economic Literature* 17: 503–22.
- Boudreaux, Donald J., and Randall G. Holcombe. 1989. "The Coasian and Knightian Theories of the Firm." *Managerial and Decision Economics* 10: 147–54.
- Caldwell, Bruce J. 1980. "A Critique of Friedman's Methodological Instrumentalism." *Southern Economic Journal* 47 (2): 366–74.
- . 1984. *Appraisal and Criticism in Economics: A Book of Readings*. Boston: Allen and Unwin.

- Coase, Ronald H. 1937. "The Nature of the Firm." In Coase, 1988b. Pp. 33–55.
- . 1972. "Industrial Organization: A Proposal for Research." In Coase, 1988b. Pp. 57–74.
- . 1988a. "The Firm, the Market, and the Law." In Coase, 1988b. Pp. 1–31.
- . 1988b. *The Firm, the Market and the Law*. Chicago: University of Chicago Press.
- . 1991. "The Institutional Structure of Production." *American Economic Review* 82: 713–19.
- Conner, Kathleen. 1991. "An Historical Comparison of Resource-Based Theory and Five Schools of Thought within Industrial Organization Economics: Do We Have a New Theory of the Firm?" *Journal of Management* 17: 121–54.
- De Marchi, Neil, ed. 1988. *The Popperian Legacy in Economics*. Cambridge: Cambridge University Press.
- Dickinson, Henry Douglas. 1933. "Price Formation in a Socialist Community." *Economic Journal* 43: 237–50.
- Eccles, Robert, and Harrison White. 1988. "Price and Authority in Inter-Profiter Transactions." *American Journal of Sociology* 94: Supplement, S17–S48.
- Ellig, Jerry. 1993. "Internal Pricing for Corporate Services." Working Paper. Center for the Study of Market Processes, George Mason University.
- Ellig, Jerry, and Wayne Gable. 1993. *Introduction to Market-Based Management*. Fairfax, Va.: Center for Market Processes.
- Ericson, Richard E. 1991. "The Classical Soviet-Type Economy: Nature of the System and Implications for Reform." *Journal of Economic Perspectives* 5 (4): 11–27.
- Fama, Eugene F. 1980. "Agency Problems and the Theory of the Firm." *Journal of Political Economy* 88: 288–307.
- Foss, Nicolai Juul. 1993a. "More on Knight and the Theory of the Firm." *Managerial and Decision Economics* 14: 269–76.
- . 1993b. "Theories of the Firm: Contractual and Competence Perspectives." *Journal of Evolutionary Economics* 3: 127–44.
- . 1993c. "The Two Coasian Traditions." *Review of Political Economy* 4: 508–32.
- . 1994. "The Theory of the Firm: The Austrians as Precursors and Critics of Contemporary Theory." *Review of Austrian Economics* 7 (1): 31–65.
- Furubotn, Eirik, and Rudolf Richter. 1991. *The New Institutional Economics: An Assessment*. College Station, Texas: Texas A&M Press.
- Gilson, Ronald J. 1995. "Corporate Governance and Economic Efficiency: When Do Institutions Matter?" Working Paper no. 121, John M. Olin Program in Law and Economics, Stanford Law School.
- Gray, Alexander. 1946. *The Socialist Tradition: Moses to Lenin*. London: Longmans, Green.
- Grossman, Sanford J., and Oliver D. Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy* 94: 691–719.
- Hayek, F. A., ed. 1935a. *Collectivist Economic Planning*. Clifton, N.J.: Augustus M. Kelley, 1975.
- . 1935b. *Prices and Production*. London: Routledge and Kegan Paul.
- . 1937. "Economics and Knowledge." In Hayek, 1948. Pp. 33–56.
- . 1940. "The Competitive Solution." In Hayek, 1948. Pp. 181–208.

- . 1945. "The Use of Knowledge in Society." In Hayek, 1948. Pp. 77–91.
- . 1948. *Individualism and Economic Order*. Chicago: University of Chicago Press.
- Heilbroner, Robert L. 1970. *Between Capitalism and Socialism: Essays in Political Economics*. New York: Random House.
- . 1990. "Reflections After Communism." *The New Yorker* (September 10): 91–100.
- Hirshleifer, Jack. 1956. "On the Economics of Transfer Pricing." *Journal of Business* 29: 172–89.
- Hoff, T. J. B. 1949. *Economic Calculation in the Socialist Society*. Indianapolis: Liberty Press, 1981.
- Holmström, Bengt R., and Jean Tirole. 1989. "The Theory of the Firm." In Richard Schmalensee and Robert D. Willig, eds., *Handbook of Industrial Organization*. Vol. 1. Amsterdam: North-Holland. Pp. 61–133.
- Jacobson, Robert. 1992. "The 'Austrian' School of Strategy." *Academy of Management Review* 17 (4): 782–807.
- King, Elizabeth A. 1994. *Transfer Pricing and Valuation in Corporate Taxation: Federal Legislation vs. Administrative Practice*. Boston: Kluwer Academic Publishers.
- Kirzner, Israel M. 1988. "The Economic Calculation Debate: Lessons for Austrians." *Review of Austrian Economics* 2: 1–18.
- Klein, Benjamin, Robert A. Crawford, and Armen A. Alchian. 1978. "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process." *Journal of Law and Economics* 21: 297–326.
- Klein, Peter G. 1994. "Mergers and the Market for Corporate Control." In Boettke, ed., 1994. Pp. 394–401.
- Knight, Frank H. 1921. *Risk, Uncertainty, and Profit*. New York: Hart, Schaffner and Marx.
- Kornai, Janos. 1986. "The Soft Budget Constraint." *Kyklos* 39 (1): 3–30.
- Kreps, David M. 1990. *A Course in Microeconomic Theory*. Princeton, N.J.: Princeton University Press.
- Lange, Oskar. 1936–37. "On the Economic Theory of Socialism." *Review of Economic Studies* 4 (October 1936): 53–71, and 4 (February 1937): 132–42. Revised and republished in Lippincott, 1938. Pp. 55–143.
- . 1958. "The Role of Planning in Socialist Economy." In Morris Bornstein, ed. *Comparative Economic Systems*. Homewood, Ill.: Richard D. Irwin, 1969. Pp. 169–81.
- . 1965. "The Computer and the Market." In Alec Nove and D. M. Nuti, eds., *Socialist Economics*. London: Penguin Books, 1972. Pp. 401–5.
- Langlois, Richard N. 1988. "Economic Change and the Boundaries of the Firm." *Journal of Institutional and Theoretical Economics* 144 (4): 635–57.
- . 1992a. "Orders and Organizations: Toward an Austrian Theory of Social Institutions." In Bruce J. Caldwell and Stephan Boehm, eds., *Austrian Economics: Tensions and New Directions*. Boston, Dordrecht, and London: Kluwer Academic Publishers. Pp. 165–83.
- . 1992b. "Transaction Cost Economics in Real Time." *Industrial and Corporate Change* 1 (1): 99–127.
- . 1994a. "The Boundaries of the Firm." In Boettke, ed., 1994. Pp. 173–78.
- . 1994b. "The 'New' Institutional Economics." In Boettke, ed., 1994. Pp. 535–40.

- Lavoie, Don. 1985. *Rivalry and Central Planning: The Socialist Calculation Debate Reconsidered*. Cambridge: Cambridge University Press.
- Lerner, Abba. 1934. "Economic Theory and Socialist Economy." *Review of Economic Studies* 2: 51–61.
- Lippincott, Benjamin E., ed. 1938. *On the Economic Theory of Socialism*. New York: McGraw-Hill, 1964.
- Machlup, Fritz. 1934. *Führer durch die Krisenpolitik*. Vienna: J. Springer.
- . 1974. "Integrationshemmende Integrationspolitik." In Herbert Giersch, ed., *Berhnard-Harms-Vorlesungen*. Kiel: Institut für Weltwirtschaft.
- . 1976. "Discussion of the Four Papers." In Moss, ed., 1976. Pp. 111–15.
- Manne, Henry G. 1965. "Mergers and the Market for Corporate Control." *Journal of Political Economy* 73 (April): 110–20.
- Milgrom, Paul A., and John Roberts. 1992. *Economics, Organization, and Management*. Englewood Cliffs, N.J.: Prentice-Hall.
- Minkler, Alanson P. 1993a. "Knowledge and Internal Organization." *Journal of Economic Behavior and Organization* 21 (1): 17–30.
- . 1993b. "The Problem with Dispersed Knowledge: Firms in Theory and Practice." *Kyklos* 46 (4): 569–87.
- Mises, Ludwig von. 1920. "Economic Calculation in the Socialist Commonwealth." Trans. S. Adler. Pp. 87–130 in Hayek, ed., 1935a.
- . 1944. *Bureaucracy*. New Haven: Yale University Press.
- . 1947. *Planned Chaos*. Irvington-on-Hudson, N.Y.: Foundation for Economic Education. Reprinted as "Epilogue" to Mises, *Socialism: An Economic and Sociological Analysis*. Indianapolis: LibertyClassics, 1981.
- . 1949. *Human Action: A Treatise on Economics*. 3rd. rev. ed. Chicago: Henry Regnery, 1966.
- Moss, Laurence S., ed. 1976. *The Economics of Ludwig von Mises: Toward a Critical Reappraisal*. Kansas City: Sheed and Ward.
- Musgrave, Alan. 1981. "'Unreal Assumptions' in Economic Theory: The F-Twist Untwisted." *Kyklos* 34: 377–87.
- Nelson, Richard R. 1991. "Why Do Firms Differ, and How Does It Matter?" *Strategic Management Journal* 12: 61–74.
- Penrose, Edith. 1959. *The Theory of the Growth of the Firm*. 3rd. ed. New York: Oxford University Press, 1995.
- Pierson, N. G. 1902. "The Problem of Value in the Socialist Community." In Hayek, 1935a. Pp. 41–85.
- Putterman, Louis, ed. 1986. *The Economic Nature of the Firm: A Reader*. Cambridge: Cambridge University Press.
- Richardson, G. B. 1972. "The Organisation of Industry." *Economic Journal* 82: 883–96. Reprinted in Richardson, *Information and Investment: A Study in the Working of the Competitive Economy*. Oxford: Clarendon Press, 1990. Pp. 224–42.
- Rizzo, Mario J. 1978. "Praxeology and Econometrics: A Critique of Positivist Economics." In Louis M. Spadaro, ed., *New Directions in Austrian Economics*. Kansas City, Mo.: Sheed Andrews and McMeel. Pp. 40–56.
- Robbins, Lionel. 1934. *The Great Depression*. New York: Macmillan.

- Roe, Mark J. 1994. *Strong Managers, Weak Owners: The Political Roots of American Corporate Finance*. Princeton: Princeton University Press.
- Rogerson, William P. 1992. *Overhead Allocation and Incentives for Cost Minimization in Defense Procurement*. Santa Monica, Calif.: Rand Corporation.
- Romano, Roberta. 1992. "A Guide to Takeovers: Theory, Evidence, and Regulation." *Yale Journal on Regulation* 9: 119–80.
- Rothbard, Murray N. 1962. *Man, Economy, and State: A Treatise on Economic Principles*. 2 Vols. Los Angeles: Nash, 1970.
- . 1976. "Ludwig von Mises and Economic Calculation Under Socialism." In Moss, ed. 1976. Pp. 67–77.
- . 1991. "The End of Socialism and the Calculation Debate Revisited." *Review of Austrian Economics* 5 (2): 51–76.
- . 1993. Letter to Sandra K. Johnson, January 20.
- Rubin, Paul H. 1990. *Managing Business Transactions: Controlling the Cost of Coordinating, Communicating, and Decision Making*. New York: Macmillan.
- Salerno, Joseph T. 1990. "Postscript" to reprint of Mises, 1920. Auburn, Ala.: Ludwig von Mises Institute.
- Schumpeter, Joseph A. 1942. *Capitalism, Socialism and Democracy*. New York: Harper and Row.
- Shelanski, Howard A. 1993. *Transfer Pricing and the Organization of Intrafirm Exchange*. Unpublished Ph.D. Dissertation, Department of Economics, University of California, Berkeley.
- Shelanski, Howard A., and Peter G. Klein. 1995. "Empirical Research in Transaction Cost Economics: A Review and Assessment." *Journal of Law, Economics and Organization* 11 (2): 335–61.
- Staubus, George J. 1986. "The Market Simulation Theory of Accounting Measurement." *Accounting and Business Research* 16 (62): 117–32.
- Taylor, Fred M. 1929. "The Guidance of Production in a Socialist State." In Lippincott, 1938. Pp. 41–54.
- Teece, David J. 1980. "Economies of Scope and the Scope of the Enterprise." *Journal of Economic Behavior and Organization* 1 (3): 223–47.
- . 1982. "Towards an Economic Theory of the Multi-Product Firm." *Journal of Economic Behavior and Organization* 3: 39–64.
- Teece, David J., Gary Pisano, and Amy Shuen. 1992. "Dynamic Capabilities and Strategic Management." Working Paper. Center for Research in Management. Haas School of Business. University of California at Berkeley.
- Weber, Max. 1921. *Economy and Society*. Berkeley: University of California Press, 1978.
- Williamson, Oliver E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press.
- . *The Economic Institutions of Capitalism*. New York: Free Press.
- . 1991. "Private Ownership and the Capital Market." In Horst Siebert, ed., *Privatization: Symposium in Honor of Herbert Giersch*. Tubingen: J.C.B. Mohr (Paul Siebeck). Pp. 27–54.
- . 1996. *The Mechanisms of Governance*. New York: Oxford University Press.
- Winiecki, Jan. 1990. "Why Economic Reforms Fail in the Soviet Union." *Economic Inquiry* 28: 195–221.

Cartels as Efficient Productive Structures

Pascal Salin

Cartels are considered to be specific productive structures which allow producers to exert a monopoly power. Evaluation of the working of cartels is thus closely linked to the theory of competition and monopoly which one adopts. This field of economics is certainly one of those in which Murray Rothbard¹ has made breakthrough contributions. He has persuasively shown that there is no monopoly power as far as there are only voluntary arrangements. From this point of view, freedom of entry in production is the only relevant criterion to evaluate productive structures and one might dismiss as irrelevant all the other traditional conditions of the pure and perfect competition theory. In the present article we will not address this general debate about competition and monopoly and we accept the views of Murray Rothbard without any discussion. Our precise aim is rather to look for the specific characteristics of cartels and to evaluate them under the light of this approach. Cartels are generally considered negatively as formal arrangements to restrict production. After having discussed this approach we explain why cartels rather play a positive role in meeting some specific demands of the market. As a consequence they modify the frontier between the firm and the market.

The Cartel as a Restrictive Structure of Production

The traditional theory of pure and perfect competition focuses on the number of participants and, as such, introduces a strong opposition between the competitive case in which demand is perfectly elastic (from the point of view of the individual producer), and the monopoly case in which demand is inelastic, which makes possible for the monopolist to extract a “super-profit.” In between is the case of the

*Pascal Salin is professor of economics at the Université Paris–Dauphiné.

¹Particularly in *Man, Economy, and State*, 3rd ed. (Auburn, Ala.: Ludwig von Mises Institute, 1993). In the present paper we rely mainly on this book, from which we extracted all quotations for which pages are given without any further indication.

oligopoly, namely the troublesome case in which there are a few producers on a market.²

What is wrong in the traditional theory is not the *formal* analysis of monopoly or oligopoly, but the confusion it has made between the end result and the process. The number of firms on a specific market at one given time is the provisional outcome of a specific process and it has no meaning by itself. On the contrary, the fact that the process is based on the use of public constraint or is purely spontaneous makes the difference. Thus, it is not true that super-profit exists whenever there is freedom of entry on a market, since one cannot explain, in such a case, why there are no more producers trying to get such a super-profit. But the traditional analysis of monopoly is clearly and directly appropriate when applied to the case of a public privilege, namely the existence of a public barrier to entry: In that latter case, the formal analysis of monopoly is a correct description of reality. In this case only does a super-profit exist.

Now, what about the case of a small number of producers, which is the scope of analysis both of the theory of oligopoly and the theory of cartels? Let us assume that for some technical or historical reasons, only a limited number of firms exist at one given time (the possible economies of scale not being large enough as to justify the existence of only one producer). Each producer has to elaborate a strategy of his own in an environment of imperfect information (since the decisions of each producer affect the state of the market, but no one knows about the intentions of others). Thus, authors such as Heinrich von Stackelberg have shown that the outcome of the process may be different according to the strategies followed by different producers. There are, for instance, cooperative and non-cooperative processes, equality between the producers or a distinction between a leader and followers, etc. Once more, these assumptions may have interesting applications. They are debatable as far as they tend to introduce the idea of a monopoly situation in the absence of any public privilege.

A distinction seems to be frequently made between an oligopoly and a cartel. A cartel could be defined as a specific oligopoly in which the different producers, instead of just interacting, enter into a process of explicit cooperation. Thus, a cartel is viewed as an agreement between different producers in order to follow common rules or behaviors, i.e., a system of mutual and freely accepted obligations.

²The definition of a market is arbitrary, since it implies the possibility to differentiate perfectly specific goods. This would be possible without any ambiguity only if goods were perfectly not substitutable one to the other. As far as some substitutability exists—which is always the case—the definition of a good and of a market is arbitrary.

It is usually assumed that this agreement can be explained by a common *intention* of participants, namely creating a monopoly position. Therefore, participants who do not face a perfectly elastic demand are able to extract a super-profit from purchasers, which they would be unable to obtain without such an agreement. Thus, cartels are mainly viewed as having a negative aspect, as is expressed in the usual saying that cartels are collusive.

In the traditional theory the super-profit is shared between participants according to *a priori* rules, for instance their market shares. It is well-known that cartels are considered unstable structures, since participants are tempted to become free riders: Although they have an interest in others' respecting the agreement, they can expect a larger share of the common super-profit by introducing some degree of competition with regard to their partners.

The cartel is sometimes considered as worse than the monopoly. In fact, it is often assumed that a monopoly exists for purely technical reasons, for instance because techniques are such that large economies of scale make it impossible for several firms to coexist in the most profitable way (natural monopoly).³ In such cases no normative judgment is issued against the monopoly, but it is argued only that the state has to prevent the assumed *exploitation* of consumers by regulating the monopoly or by nationalizing it. The value judgment concerning cartels is more critical since it is assumed that there is no technical reason for any monopolist position, but that the cartel is created *ex nihilo* by an explicit agreement between producers in order to exploit purchasers. From this point of view any cartel arrangement ought to be prohibited, for the cartel exists *in order to* create restrictive procedures.

In fact, as we shall see later, this negative approach to the cartel is a direct consequence of the traditional theory of competition. According to it, pure or perfect competition can only exist when a great number of (identical) producers produce a homogeneous good. In fact this theory does not describe a real process of competition between real entrepreneurs, but the technical organization of managers in a non-innovative system, for instance the managers of plants in a Soviet-style centralized system of production: It is assumed that there is one single technique—an *optimal* technique from a purely technical point of view—to produce a given good and all (numerous) managers have to apply exactly the same technique to produce exactly the same good.

³Anyhow, those who support such views do not usually consider the possibility of competition from producers located in other nations and are ready to support protectionist measures.

In that sense, it can be said that the theory of pure and perfect competition is in fact a theory of central planning.⁴

The Austrian theory of competition leads to an opposite view. The end result of a competitive process is not the production of identical goods by numerous producers, but, on the contrary, the *differentiation* of goods produced by entrepreneurs who are innovators and who are induced by competition—namely the free entry of potential competitors—to offer better goods at lower prices.

Therefore, as far as cartel agreements intend the homogenization of goods (either of their specific characteristics or their price), one may wonder for which category of theories it could be a problem. From the point of view of a supporter of the traditional approach of competition, the problem may not be that of homogenizing goods, but of creating monopoly power. From the point of view of the supporter of the Austrian theory of competition, there is no such thing as monopoly power, as far as the process of monopolization by cartels is spontaneous. But the reason for homogenization has to be clarified—which we will try to do later—since the merit of competition relies on the fact that it induces *differentiation*.

The approach used by Murray Rothbard to study the specific case of cartels seems to be the following: He accepts the idea according to which cartels are a system of collusion aiming at introducing some restrictive actions in a productive system (rarefaction of production). However, he shows that, on the one hand, restrictive actions of producers are perfectly justified and that, on the other hand, cartels cannot create monopoly power since such power cannot exist as long as there is free entry on a market. Thus, in Murray Rothbard's view, if producers organize themselves into a restrictive agreement—a cartel—it is to meet a demand by purchasers for a restrictive action. The argument goes as follows. A monopoly position can be reached by a single producer or a set of producers organized into a cartel only if demand is inelastic, which makes restriction of production possible and profitable. But, it is up to demanders not to have an inelastic demand: If ever they were unsatisfied with the cartel's behavior, they could decide on having a perfectly elastic demand, i.e., not accepting an increase in prices through the restriction of production. As Murray Rothbard wrote,

If the consumers were really opposed to the cartel action, and if the resulting exchanges really hurt them, they would boycott the

⁴We have developed this view in a short book, *La concurrence*, in the *Que sais-je?* collection (Paris: Presses Universitaires de France, 1995).

“monopolistic” firm or firms, they would lower their purchasing so that the demand curve became *elastic*, and the firm would be forced to increase its production and reduce its price again. If the “monopolistic price” action had been taken by a cartel of firms, and the cartel had no other advantages for rendering production more efficient, it would then have to disband, because of the now demonstrated elasticity of the demand schedule.⁵

By saying that demanders could react against the cartel by lowering their purchases, he implicitly assumes that the good produced by the cartel has close substitutes, so that the demanders can decline the restrictive action of producers (and the corresponding increase in prices) by shifting to these substitutes, without any significant loss of utility.

Thus, if a cartel burns coffee in order to increase its price, the waste comes from the fact that there has been an excessive production of coffee and not from the fact it is burnt (since coffee has no other value than the one it gets from its relative scarcity in comparison with other goods). In some sense the restrictive action is not specific to a cartel, since on any free market one always restricts the production of a good in order to produce another one. According to Murray Rothbard: “The cartel’s action, in reducing the production of coffee and causing an increase in the production of rubber, jungle guiding, etc., led to an increase in the power of the productive resources to satisfy consumer desires,”⁶ and “A cartel action, if it is a voluntary one, cannot injure freedom of competition and, if it proves profitable, benefits rather than injures the consumers.”⁷ Thus, either the cartel results from the free working of the market and its possible restrictive actions are *wanted* by consumers, or it is the consequence of a barrier to entry and the restrictive action is a pure monopoly—i.e., state created—exploitation of purchasers.

However, Murray Rothbard seems to limit justification for cartels to the case of voluntary and beneficial restrictions of production, although he does not exclude the possibility of other cartel actions when he writes:

To regard a cartel as immoral or as hampering some sort of consumer’s sovereignty is therefore completely unwarranted. And this is true even in the seemingly “worst” case of a cartel that we may assume is founded *solely* for “restrictive” purposes, and where, as a

⁵Murray Rothbard, *Man, Economy, and State*, p. 565.

⁶*Ibid.*, p. 569.

⁷*Ibid.*, p. 584.

result of previous error and the perishability of product, actual destruction will occur.⁸

However, the very notion of a "restrictive action" is by itself debatable, inasmuch as such an action can be labeled as restrictive only in comparison with a norm which could undoubtedly be named non-restrictive. Let us take the example of coffee producers burning coffee in order to increase its price and obtain a maximum profit. A cooperative action is necessary in order to avoid free-riding. However, instead of burning coffee after it has been produced, producers could have decided on a cooperative action before producing it in order to avoid overproduction. They did not do it just because they lacked the necessary information on the future state of the market. Therefore, burning coffee is a restrictive action only in comparison with a specific state of affairs which had been freely chosen by producers in a situation of imperfect information.⁹ Therefore, it would be preferable to avoid using the normative word of "restriction" and to just assume that the cartel aims at profit maximization.

Now, if there is free entry on the market, it is wrong to say that this profit maximization corresponds to a super-profit. As is the notion of a restrictive action, the notion of a super-profit is meaningless, since it implies a comparison with a completely non-existing norm, namely the rate of profit obtained in a situation of pure or perfect competition.

The mere definition of a cartel does not imply anything about the degree of freedom of entry on the market. The theory of monopoly has no scope as long as the reasons for the existence of one single producer are not made precise (e.g., barriers to entry, economies of scale or innovation). Similarly, the theory of cartels is meaningless without a clear understanding of the reasons for the existence of such a market structure. As we have seen, the possibility to introduce a restrictive action on a market is not an acceptable explanation and we have to consider other justifications for cartels.

If there is not a single producer but a couple of producers in a given market, it means for instance that there is no additional gain to be obtained from shifting from several producers to one single producer (the optimal size of firms has been obtained with several firms and there is no marginal economy of scale). To most theorists, cartels appear as unstable market structures, because they can find explanations for a situation with one monopolist or for a situation with a great number

⁸Ibid., p. 570.

⁹Moreover, in such a case, the cartel is only a transitory market structure to cope with the unexpected consequences of non-cooperative behaviors. Now, in principle, the market is a coordination process which makes cooperation unnecessary.

of producers, but not for any in-between situation. The cartel thus appears as a transitory solution to solve a specific problem. However, this intermediate market structure can be perfectly stable and optimal.

The Cartel as a Value-Producing Organization

As we just stressed, contrary to the traditional theory, competition has the merit of inducing producers to differentiate their production one from the other and not to try to produce exactly the same good with the same technique. Now, if we define a cartel as a structure which allows different producers to coordinate their production in order to suppress any differentiation in their products, does it mean that cartels can be viewed as anti-competitive organizations, or that some other reason might explain the emergence of such a productive structure?

There is a very general and simple answer to this question: If competition prevails in the sense that there is free entry on a market and if a cartel has existed for long, it necessarily means that this structure is the best one to meet some specific demand of the market. In other words, a cartel is not necessarily unstable—as is obvious from practical experiences—and if it remains alive in spite of potential competition, it means that it is a useful structure in this specific case. In other words, competition usually leads to differentiation of products, i.e., imperfect substitution between them, whereas a cartel tends to induce homogenization, i.e., substitutability. The benefits of differentiation are so obvious that there must be serious reasons for limiting it and introducing a higher degree of substitutability between products.

In fact there are many specific activities in which there is a demand for homogeneity, especially in network activities, for instance telecommunications, transportation, or money production (which are frequently considered as public utilities and, even, natural monopolies). Generally speaking such situations can be named situations of “sub-additivity,”¹⁰ which includes externalities, economies of scale and economies of scope. In such cases one can obtain gains from coordinating several producers or from substituting one unique producer for a number of them.

As an example, in the case of money production it can be considered that there are economies of scale and, therefore, decreasing marginal costs in the production of money, since, for instance, advertising costs on the characteristics of a currency may be more or less fixed and the centralization of reserves allows savings of resources; there are

¹⁰This term seems to have been used first by the economists of Bell Co. It is used, for instance, in W. Baumol, J. Panzar, J. C. Willig, *Contestable Markets and the Theory of Industry Structure* (New York: Harcourt, Brace, Jovanovich, 1982).

economies of scope (since information obtained by a financial intermediary can be efficiently used to create money against credit); and, finally, there may be externalities, since, for instance, a currency is more useful for one person the more widely it is used by others. Even if one disagrees with the precise reasons for sub-additivity, anyone may accept the idea that it would not be *optimal* to have a very large number of different currencies.

Governments and all the *experts* who support them usually shift from such observations to the conclusion that there are natural monopolies, so that public monopolies or regulations are necessary in order to avoid the exploitation of demanders by producers. In fact, there is only one conclusion to be drawn from such observations, namely that there is a potential gain to be obtained from decreasing the degree of differentiation in the production of such goods. There are potential gains which can be obtained from substituting one or a limited number of goods for a great diversity. However, it does not mean that the optimal degree of diversification—or, conversely, the optimal degree of homogenization—can be decided *a priori* from a purely technical approach. Two remarks are important at this point:

(1) If ever gains can be obtained by diminishing the degree of differentiation—for instance because of economies of scale—it does not imply that it is also optimal to diminish the number of producers. As an example, if it could be proven that it would be technically optimal to have one single money or one single computer standard in the world, it would not imply that there ought to be one single producer: Either a monopoly or a cartel can do the job. In fact, a cartel is a productive structure in which *different* producers produce the same good, so that they can be as efficient as a monopoly in meeting sub-additivity problems.

(2) One cannot know in advance and forever whether there is sub-additivity in an activity. It has to be discovered. And, as is well-known, competition—i.e., free entry—is the best way to discover to which extent sub-additivity does exist and to which extent it may change over time. More specifically, it cannot be generally said that sub-additivity exists in such and such activity, as it may exist at one given time at one point of the production process, but not some particular over its entire range. For instance in a telecommunication or transportation network, it may not be efficient to have more than one major *highway* in some part of the network, whereas in other parts there would not be any economies of scale: the *highway* may be operated by one *monopolistic* producer or by several producers coordinated into a *cartel*, whereas the other parts could be managed by different, not coordinated, producers as well as by a cartel (for

instance the one which is managing the *highway*). As new technologies are discovered, the place of the cartel may change over time.

Therefore, cartels exist not only, or even not mainly, in order to make resources scarcer and to increase prices, but to increase the value of production and improve the productive processes. The cartel is not created to extract what the traditional theory calls a "super-profit," but to produce coordination gains.

Cartels, as well as monopolies, are the possible outcome of competition, i.e., free entry on a market: Competition makes it possible for real entrepreneurs to innovate, so that they are the only suppliers of the new product they have decided to introduce to the market. As competition has the merit of inducing producers to be the first ones on a market—i.e., to be what traditional theory calls monopolists—it is meaningless to compare cartel situations to a situation with many producers. But it makes sense to compare cartels to monopolies. In fact, if ever there are gains to be obtained from homogenizing production, because of sub-additivity phenomena, is it not more efficient to have one single producer than several ones in a cartel?

In fact, in a cartel there are possible coordination costs which may be opposite to the coordination gains obtained through the homogenization of production. However the cartel may be organized along two different procedures, either spontaneous coordination or explicit cooperation. As an example of the first category, let us take the case of a free banking system and let us assume that, in a given area, several banks produce currencies which benefit from a convertibility guarantee in terms of gold. However, each bank discovers that, in order to make its own money more attractive, it has to increase its liquidity and, therefore, it decides to accept the currencies issued by other banks of the system against its own without limit, at the fixed price given by their mutual gold prices. Such a system of spontaneous coordination makes all currencies perfectly substitutable, which means that the banks of the system have decided to eliminate any possible differentiation of their products. However, one may imagine that some degree of cooperation—i.e., explicit coordination of decisions—may take place, for instance to decide a common name for the *common* currency, but it is not absolutely necessary, since each bank can decide separately to use the name of a currency already issued by one of them. Thus, one cannot define a cartel by the existence of cooperation—or what is sometimes called collusion—but rather by the fact that there is homogenization of goods produced by different producers, whether it results from explicit, centralized decisions or from decentralized, individual decisions.

Coordination costs are certainly higher the more cooperation exists in the system, since it is necessary to monitor the explicit cartel agreement. In such a case, as we have already recalled, each participant in the cartel has an interest in others' honoring the agreement, and in trying to engage in free riding in order to get a larger share of the common market. But coordination costs are almost non-existent whenever the cartel is the result of the spontaneous decisions of its members.

On one hand, a cartel may have some advantages in comparison with the case of a single producer (monopolist). The main gain stems from the fact that, although there is actual homogenization of production, the possibility of *future diversification* remains. As we have already stressed, the existence of coordination gains may change over time as new technologies are discovered. When a single good is produced by several producers in a cartel rather than by a single producer, there may be more incentives to discover new techniques. The traditional assertion according to which a cartel is necessarily unstable is not completely wrong, but it must not be considered as a negative aspect of the cartel, but rather as a positive one. It means that the cartel subsists as long as it is the most efficient productive organization. But that it may burst whenever other productive structures appear as more efficient. Moreover, the participants in the cartel are permanently induced to look for the possibility of inventing more efficient productive organizations, some of which includes the dissolving of the cartel or, maybe, the arrangement of a new form of cartel.

Another possible gain brought by the cartel is a scale gain. Contrary to what is usually assumed economies of scale do not generally exist. Moreover, whenever it can be assumed, that there are economies of scale, they have to be viewed as *technical* and not as *institutional* economies of scale.¹¹ But there are also diseconomies of scale, which are mainly of an institutional nature. It is well-known that the internal organization of a firm does not rely on explicit exchange procedures, so that the production of information may be more difficult. The larger the firm the larger may be the organization cost. Thus, by coordinating their production—or cooperating—in a cartel, different firms are able to produce a good at the optimal scale, from the technical point of view, and of minimizing institutional costs (diseconomies of scale).

The IATA (International Air Transport Association) gives an interesting example of an efficient, rather stable, but changing cartel. Through

¹¹As Murray Rothbard rightly wrote, "The critical problem is not the size of the plant, but the size of the firm" (*Man, Economy, and State*, p. 577) and "Economics can make few valid statements about the optimal size of a firm except that the free market will come as close as possible to rendering maximum service to consumers, whether we are considering the size of a firm or any other aspect of production" (*Ibid.*, p. 578).

an agreement which is not very costly to manage, different firms are able to give more value to their services, since the tickets issued by different firms are (nearly perfectly) substitutable one for another, at least as regards regular tickets. However, airlines are allowed some degree of differentiation—for instance in the quality of service—and, on the other hand, they also produce services (e.g., special rates and charters) which do not enter the cartel agreement. Thus, the airline industry is characterized by an optimal degree of differentiation and homogenization from both the points of view of travelers and airlines. To some travelers, the substitutability between airline tickets is very valuable, for some other ones the priority consists in getting the lowest possible prices. Thus, the airline industry meets the different needs of customers.

The traditional theory of competition, as well as the traditional theory of the natural monopoly, have a global view of what they call a *good*. It is considered, for example, that there is something such as *air transportation* or *telecommunications*. Moreover it is assumed that sub-additivity exists in such network activities, so that there is the possibility of a natural monopoly. In fact, as it has been made clear by the *new* theory of consumption,¹² people do not demand *goods*, but characteristics and physically distinct goods are supplying bundles of characteristics, in various proportions. The problem of production is to adjust to the immense variety of characteristics desired by different individuals. Now, homogeneity may be one valuable—and, therefore, demanded—characteristic, as is the case for money or air transportation. Therefore, to be efficient a productive structure has to produce baskets of characteristics which are viewed as optimal by demanders, which implies that some baskets may include homogenized goods and some others differentiated ones. The cartel—which is a mix of differentiation and homogenization—contributes to this adjustment. From this point of view it is an essential feature of productive structures. Contrary to the usual view according to which cartels are fundamentally unstable and, therefore, transitory, and contrary, even, to the view of Murray Rothbard who also considered cartels as somewhat transitory organizations, cartels have a durable, although changing, role to play.

The Cartel as an Intermediate Productive Structure

Thus, the cartel plays an important role in allowing an optimal combination of diversification and homogenization in production, according

¹²K. T. Lancaster, "A New Approach to Consumer Theory," *Journal of Political Economy* (April 1966); reprinted in *Modern Consumer Theory* (Brookfield, Vt. and Aldershot, England: Edward Elgar, 1991).

to the needs of demanders; and in providing an optimal combination of coordination and cooperation. From this latter point of view, it plays an interesting and intermediate role in productive structures.

Murray Rothbard considers that a cartel aims at organizing a “co-operation to increase the incomes of the producers” and he adds: “For what is the essence of a cartel action? Individual producers agree to pool their assets into a common lot, this single central organization agrees to make the decisions on production and price policies for all the owners and then to allocate the monetary gains among them. *But is this process not the same as any sort of joint partnership or the formation of a single corporation?*”¹³ Comparing the creation of a cartel to that of a centralized firm, he concludes, “*there is therefore no essential difference between a cartel and an ordinary corporation or partnership.*” However, as we already stressed, if ever a cartel and a big firm are exactly the same, why would cartels exist? Seeing no basic difference between them, Rothbard assumes that cartels are mainly transitory structures, contrary to what can be shown by reasoning or experience.¹⁴ In fact, his conclusions are dependent on his definition of the cartel. Viewing the cartel as an explicit cooperation of firms in order to increase the joint profit—according to the traditional definition of a cartel—he assumes that the most efficient firms will be tempted to break the cartel in order to increase their market shares. In fact, as we have seen, a cartel cannot be defined by a strategy of market sharing, although it does exist, but by a coordinated—not necessarily cooperative—effort to homogenize production (which may imply identical prices). If this homogenization is desired by the market, the cartel is efficient and it will last. If it does not meet any specific need of the demanders, but is only the end result of an effort by producers to maximize joint profits, it may fail more or less rapidly.

As a consequence of his restrictive definition of a cartel, Murray Rothbard thinks that either the cartel is efficient and a merger will rapidly take place between its members, or it is not, and it will break down. As he wrote, “if joint action is the most efficient and profitable course for each member, a merger will soon take place.”¹⁵ In reality efficient cartels can and must last, possibly by transforming their structure and activities or the number of their participants. The best example may be that of money production. Under free banking the production of money by members of cartels was efficiently made without

¹³Rothbard, *Man, Economy, and State*, p. 572; emphasis in the original.

¹⁴“In many cases, a cartel can be considered as simply a tentative step in the direction of permanent merger,” *ibid.*, p. 573.

¹⁵*Ibid.*, p. 579.

any destruction of the cartel or any merger into one single big producer. In fact, if there are no unlimited economies of scale, there is no reason for a merger.

Mergers do not occur precisely because in a cartel firms are independent profit centers, which makes economic calculation more efficient. Instead of viewing a cartel as a set of firms which are about to merge, it may be both more realistic and more efficient to consider it as the ultimate stage of a process by which a big firm has been decentralized into various coordinated *decision centers* and, ultimately, split into independent *profit centers* with different owners.

As is well-known in his seminal work on the modern theory of the firm, Ronald Coase¹⁶ gave an answer to the following question: If, as it is rightly assumed, the market is an efficient way for individuals to organize their mutual exchanges, why substitute other procedures, for instance the cooperative and command procedures which are used inside a firm? The market makes it possible to *coordinate* relations between individuals through voluntary exchange, whereas individual actions are made compatible inside a firm (or any other organization) through *cooperation*, i.e., a complex mix of spontaneous and constrained (command) actions.

Therefore, it is now widely admitted that an optimal organization of production stems from the juxtaposition of two non-excluding schemes, cooperation and coordination. The market is a *coordination* process between voluntary actors and the firm—as well as any other organization—is a *cooperative* system in which the productive process is based not on spontaneous interactions between individuals through contracts but through more vague processes of cooperation (for instance through command, although in any firm there is a mix of command, voluntary decisions and initiatives, coordination through information processes, etc.). Now, there is a sort of frontier between the market process and the organizational process of firms. It can be assumed that the larger a firm is the more difficult is the internal organization process. However, the firm may gain from either the possibility of developing economies of scale or—according to the traditional theory—exploiting a monopoly position and a super-profit. From these conflicting tendencies, an optimal size of the firm results under specific conditions.

Instead of this one-or-the-other approach, a cartel makes it possible to better combine these conflicting tendencies via a better use of both cooperative solutions (internal organization) and coordination

¹⁶Ronald H. Coase, "The Nature of the Firm," *Economica* (1932).

processes (market). This is the reason why there is a great variety of cartels combining, in different ways, coordination and cooperation procedures. This is also the reason why cartels are not necessarily unstable. They are part of a firm's strategies: If ever a firm considers any change in its production, it does not have to choose only between extending the internal process of cooperation or entering into voluntary exchange on the market. It may be preferable to combine cooperation and coordination under the form of a cartel.¹⁷ The micro-computer industry gives an interesting example of such strategies. According to the evolution of markets, strategies, and technologies, producers decide their standards independently or enter into cooperative agreements which can be considered as cartel arrangements.

From this point of view it is preferable to abandon the definition of a cartel as an agreement between firms which *intends* to exert a restrictive action or any sort of specific action. The actual intention of participants is not relevant. Any action results from an intention, but the content of the intention does not matter from the market point of view. It may be that an entrepreneur enters into an agreement with some specific intention, but the outcome of the agreement is not the one intended, but another one which appears as beneficial, so that the agreement will be maintained. What is important in a cartel is that some mix of coordination-cooperation efficiently blurs the frontier between organizational processes and market processes.

¹⁷Franchising in retail trade is a well-known example of such strategies.

The Myth of Natural Monopoly

Thomas J. DiLorenzo

The very term “public utility” . . . is an absurd one. *Every* good is useful “to the public,” and almost every good . . . may be considered “necessary.” Any designation of a few industries as “public utilities” is completely arbitrary and unjustified.

—Murray Rothbard, *Power and Market*

Most so-called public utilities have been granted governmental franchise monopolies because they are thought to be “natural monopolies.” Put simply, a natural monopoly is said to occur when production technology, such as relatively high fixed costs, causes long-run average total costs to decline as output expands. In such industries, the theory goes, a single producer will eventually be able to produce at a lower cost than any two other producers, thereby creating a “natural” monopoly. Higher prices will result if more than one producer supplies the market.

Furthermore, competition is said to cause consumer inconvenience because of the construction of duplicative facilities, e.g., digging up the streets to put in dual gas or water lines. Avoiding such inconveniences is another reason offered for government franchise monopolies for industries with declining long-run average total costs.

It is a myth that natural monopoly theory was developed first by economists, and then used by legislators to “justify” franchise monopolies. The truth is that the monopolies were created decades before the theory was formalized by intervention-minded economists, who then used the theory as an *ex post* rationale for government intervention. At the time when the first government franchise monopolies were being granted, the large majority of economists understood that large-scale, capital intensive production did *not* lead to monopoly, but was an absolutely desirable aspect of the competitive process.

*Thomas J. DiLorenzo is professor of economics at the Sellinger School of Business and Management, Loyola College.

The Review of Austrian Economics Vol. 9, No. 2 (1996): 43–58
ISSN 0889-3047

The word “process” is important here. If competition is viewed as a dynamic, rivalrous process of entrepreneurship, then the fact that a single producer happens to have the lowest costs *at any one point in time* is of little or no consequence. The enduring forces of competition—including potential competition—will render free-market monopoly an impossibility.

The theory of natural monopoly is also a-historical. There is no evidence of the “natural monopoly” story ever having been carried out—of one producer achieving lower long-run average total costs than everyone else in the industry and thereby establishing a permanent monopoly. As discussed below, in many of the so-called public utility industries of the late eighteenth and early nineteenth centuries, there were often literally dozens of competitors.

Economies of Scale During the Franchise Monopoly Era

During the late nineteenth century, when local governments were beginning to grant franchise monopolies, the general economic understanding was that “monopoly” was caused by government intervention, not the free market, through franchises, protectionism, and other means. Large-scale production and economies of scale were seen as a competitive virtue, not a monopolistic vice. For example, Richard T. Ely, co-founder of the American Economic Association, wrote that “large scale production is a thing which by no means necessarily signifies monopolized production.”¹ John Bates Clark, Ely’s co-founder, wrote in 1888 that the notion that industrial combinations would “destroy competition” should “not be too hastily accepted.”²

Herbert Davenport of the University of Chicago advised in 1919 that only a few firms in an industry where there are economies of scale does not “require the elimination of competition,”³ and his colleague, James Laughlin, noted that even when “a combination is large, a rival combination may give the most spirited competition.”⁴ Irving Fisher⁵ and Edwin R.A. Seligman⁶ both agreed that large-scale production produced

¹Richard T. Ely, *Monopolies and Trusts* (New York: MacMillan, 1990), p. 162.

²John Bates Clark and Franklin Giddings, *Modern Distributive Processes* (Boston: Ginn & Co., 1888), p. 21.

³Herbert Davenport, *The Economics of Enterprise* (New York: MacMillan, 1919), p. 483.

⁴James L. Laughlin, *The Elements of Political Economy* (New York: American Book, 1902), p. 71.

⁵Irving Fisher, *Elementary Principles of Economics* (New York: MacMillan, 1912), p. 330.

⁶E. R. A. Seligman, *Principles of Economics* (New York: Longmans, Green, 1909), p. 341.

competitive benefits through cost savings in advertising, selling, and less cross-shipping.

Large-scale production units unequivocally benefited the consumer, according to turn-of-the-century economists. For without large-scale production, according to Seligman, "the world would revert to a more primitive state of well being, and would virtually renounce the inestimable benefits of the best utilization of capital."⁷ Simon Patten of the Wharton School expressed a similar view that "the combination of capital does not cause any economic disadvantage to the community. . . . combinations are much more efficient than were the small producers whom they displaced."⁸

Like virtually every other economist of the day, Columbia's Franklin Giddings viewed competition much like the modern-day Austrian economists do, as a dynamic, rivalrous process. Consequently, he observed that "competition in some form is a permanent economic process. . . . Therefore, when market competition seems to have been suppressed, we should inquire what has become of the forces by which it was generated. We should inquire, further, to what degree market competition actually is suppressed or converted into other forms."⁹ In other words, a "dominant" firm that underprices all its rivals at any one point in time has not suppressed competition, for competition is "a permanent economic process."

David A. Wells, one of the most popular economic writers of the late nineteenth century, wrote that "the world demands abundance of commodities, and demands them cheaply; and experience shows that it can have them only by the employment of great capital upon extensive scale."¹⁰ And George Gunton believed that "concentration of capital does not drive small capitalists out of business, but simply integrates them into larger and more complex systems of production, in which they are enabled to produce . . . more cheaply for the community and obtain a larger income for themselves. . . . Instead of concentration of capital tending to destroy competition the reverse is true. . . . By the use of large capital, improved machinery and better facilities the trust can and does undersell the corporation."¹¹

The above quotations are not a selected, but rather a comprehensive list. It may seem odd by today's standards, but as A.W. Coats

⁷Ibid., p. 97.

⁸Simon Patten, "The Economic Effects of Combinations," *Age of Steel* (Jan. 5, 1889): 13.

⁹Franklin Giddings, "The Persistence of Competition," *Political Science Quarterly* (March 1887): 62.

¹⁰David A. Wells, *Recent Economic Changes* (New York: DeCapro Press, 1889), p. 74.

¹¹George Gunton, "The Economics and Social Aspects of Trusts," *Political Science Quarterly* (Sept. 1888): 385.

pointed out, by the late 1880s there were only ten men who had attained full-time professional status as economists in the U.S.¹² Thus, the above quotations cover virtually every professional economist who had anything to say about the relationship between economies of scale and competitiveness at the turn of the century.

The significance of these views is that these men observed firsthand the advent of large-scale production and did not see it leading to monopoly, “natural” or otherwise. In the spirit of the Austrian School, they understood that competition was an ongoing process, and that market dominance was always necessarily temporary in the absence of monopoly-creating government regulation. This view is also consistent with my own research findings that the “trusts” of the late nineteenth century were in fact dropping their prices and expanding output faster than the rest of the economy—they were the most dynamic and competitive of all industries, not monopolists.¹³ Perhaps this is why they were targeted by protectionist legislators and subjected to “antitrust” laws.

The economics profession came to embrace the theory of natural monopoly after the 1920s, when it became infatuated with “scientism” and adopted a more or less engineering theory of competition that categorized industries in terms of constant, decreasing, and increasing returns to scale (declining average total costs). According to this way of thinking, engineering relationships determined market structure and, consequently, competitiveness. The meaning of competition was no longer viewed as a behavioral phenomenon, but an engineering relationship. With the exception of such economists as Joseph Schumpeter, Ludwig von Mises, Friedrich Hayek, and other members of the Austrian School, the ongoing *process* of competitive rivalry and entrepreneurship was largely ignored.

How “Natural” Were the Early Natural Monopolies?

There is no evidence at all that at the outset of public utility regulation there existed any such phenomenon as a “natural monopoly.” As Harold Demsetz has pointed out:

Six electric light companies were organized in the one year of 1887 in New York City. Forty-five electric light enterprises had the legal right to operate in Chicago in 1907. Prior to 1895, Duluth, Minnesota, was served by five electric lighting companies, and Scranton, Pennsylvania, had four in 1906. . . . During the latter part of the nineteenth

¹²A. W. Coats, “The American Political Economy Club,” *American Economic Review* (Sept. 1961): 621–37.

¹³Thomas J. DiLorenzo, “The Origins of Antitrust: An Interest-Group Perspective,” *International Review of Law and Economics* (Fall 1985): 73–90.

century, competition was the usual situation in the gas industry in this country. Before 1884, six competing companies were operating in New York City . . . competition was common and especially persistent in the telephone industry . . . Baltimore, Chicago, Cleveland, Columbus, Detroit, Kansas City, Minneapolis, Philadelphia, Pittsburgh, and St. Louis, among the larger cities, had at least two telephone services in 1905.¹⁴

In an extreme understatement, Demsetz concludes that "one begins to doubt that scale economies characterized the utility industry at the time when regulation replaced market competition."¹⁵

A most instructive example of the non-existence of natural monopoly in the utility industries is provided in a 1936 book by economist George T. Brown entitled "The Gas Light Company of Baltimore," which bears the misleading subtitle, "A Study of Natural Monopoly."¹⁶ The book presents "the study of the evolutionary character of utilities" in general, with special emphasis on the Gas Light Company of Baltimore, the problems of which "are not peculiar either to the Baltimore company or the State of Maryland, but are typical of those met everywhere in the public utility industry."¹⁷

The history of the Gas Light Company of Baltimore figures prominently in the whole history of natural monopoly, in theory and in practice, for the influential Richard T. Ely, who was a professor of economics at Johns Hopkins University in Baltimore, chronicled the company's problems in a series of articles in the *Baltimore Sun* that were later published as a widely-sold book. Much of Ely's analysis came to be the accepted economic dogma with regard to the theory of natural monopoly.

The history of the Gas Light Company of Baltimore is that, from its founding in 1816, it constantly struggled with new competitors. Its response was not only to try to compete in the marketplace, but also to lobby the state and local government authorities to refrain from granting corporate charters to its competitors. The company operated with economies of scale, but that did not prevent numerous competitors from cropping up.

"Competition is the life of business," the *Baltimore Sun* editorialized in 1851 as it welcomed news of new competitors in the gas light

¹⁴Burton N. Behling, "Competition and Monopoly in Public Utility Industries" (1938), in Harold Demsetz, ed., *Efficiency, Competition, and Policy* (Cambridge, Mass.: Blackwell, 1989), p. 78.

¹⁵*Ibid.*

¹⁶George T. Brown, *The Gas Light Company of Baltimore: A Study of Natural Monopoly* (Baltimore, Maryland: Johns Hopkins University Press, 1936).

¹⁷*Ibid.*, p. 5.

business.¹⁸ The Gas Light Company of Baltimore, however, “objected to the granting of franchise rights to the new company.”¹⁹

Brown states that “gas companies in other cities were exposed to ruinous competition,” and then catalogues how those same companies sought desperately to enter the Baltimore market. But if such competition was so “ruinous,” why would these companies enter new—and presumably just as “ruinous”—markets? Either Brown’s theory of “ruinous competition”—which soon came to be the generally accepted one—was incorrect, or those companies were irrational gluttons for financial punishment.

By ignoring the *dynamic* nature of the competitive process, Brown made the same mistake that many other economists still make: believing that “excessive” competition can be “destructive” if low-cost producers drive their less efficient rivals from the market.²⁰ Such competition may be “destructive” to high-cost competitors, but it is beneficial to consumers.

In 1880 there were three competing gas companies in Baltimore who fiercely competed with one another. They tried to merge and operate as a monopolist in 1888, but a new competitor foiled their plans: “Thomas Alva Edison introduced the electric light which threatened the existence of all gas companies.”²¹ From that point on there was competition between both gas and electric companies, all of which incurred heavy fixed costs which led to economies of scale. Nevertheless, no free-market or “natural” monopoly ever materialized.

When monopoly did appear, it was solely because of government intervention. For example, in 1890 a bill was introduced into the Maryland legislature which “called for an annual payment to the city from the Consolidated [Gas Company] of \$10,000 a year and 3 percent of all dividends declared in return for the privilege of enjoying a 25-year monopoly.”²² This is the now-familiar approach of government officials colluding with industry executives to establish a monopoly that will gouge the consumers, and then sharing the loot with the politicians in the form of franchise fees and taxes on monopoly revenues. This approach is especially pervasive today in the cable TV industry.

Legislative “regulation” of gas and electric companies produced the predictable result of monopoly prices, which the public complained bitterly about. Rather than deregulating the industry and letting competition control prices, however, public utility regulation was adopted to supposedly appease the consumers who, according to Brown, “felt that

¹⁸Ibid., p. 31.

¹⁹Ibid.

²⁰Ibid., p. 47.

²¹Ibid., p. 52.

²²Ibid., p. 75.

the negligent manner in which their interests were being served [by legislative control of gas and electric prices] resulted in high rates and monopoly privileges. *The development of utility regulation in Maryland typified the experience of other states.*"²³

Not all economists were fooled by the "natural monopoly" theory advocated by utility industry monopolists and their paid economic advisers. In 1940 economist Horace M. Gray, an assistant dean of the graduate school at the University of Illinois, surveyed the history of "the public utility concept," including the theory of "natural" monopoly. "During the nineteenth century," Gray observed, it was widely believed that "the public interest would be best promoted by grants of special privilege to private persons and to corporations" in many industries.²⁴ This included patents, subsidies, tariffs, land grants to the railroads, and monopoly franchises for "public" utilities. "The final result was monopoly, exploitation, and political corruption."²⁵ With regard to "public" utilities, Gray records that "between 1907 and 1938, the policy of state-created, state-protected monopoly became firmly established over a significant portion of the economy and became the keystone of modern public utility regulation."²⁶ From that time on, "the public utility status was to be the haven of refuge for all aspiring monopolists who found it too difficult, too costly, or too precarious to secure and maintain monopoly by private action alone."²⁷

In support of this contention, Gray pointed out how virtually every aspiring monopolist in the country tried to be designated a "public utility," including the radio, real estate, milk, air transport, coal, oil, and agricultural industries, to name but a few. Along these same lines, "the whole NRA experiment may be regarded as an effort by big business to secure legal sanction for its monopolistic practices."²⁸ Those lucky industries that were able to be politically designated as "public utilities" also used the public utility concept to keep out the competition.

The role of economists in this scheme was to construct what Gray called a "confused rationalization" for "the sinister forces of private privilege and monopoly," i.e., the theory of "natural" monopoly. "The protection of consumers faded into the background."²⁹

More recent economic research supports Gray's analysis. In one of the first statistical studies of the effects of rate regulation in the electric

²³Ibid., p. 106. Emphasis added.

²⁴Horace M. Gray, "The Passing of the Public Utility Concept," *Journal of Land and Public Utility Economics* (Feb. 1940): 8.

²⁵Ibid.

²⁶Ibid., p. 9.

²⁷Ibid.

²⁸Ibid., p. 15.

²⁹Ibid., p. 11.

utilities industry, published in 1962, George Stigler and Claire Friedland found no significant differences in prices and profits of utilities with and without regulatory commissions from 1917 to 1932.³⁰ Early rate regulators *did not* benefit the consumer, but were rather “captured” by the industry, as happened in so many other industries, from trucking to airlines to cable television. It is noteworthy—but not very laudable—that it took economists almost 50 years to begin studying the actual, as opposed to the theoretical, effects of rate regulation.

Sixteen years after the Stigler–Friedland study, Gregg Jarrell observed that 25 states substituted state for municipal regulation of electric power ratemaking between 1912 and 1917, the effects of which were to *raise* prices by 46 percent and profits by 38 percent, while reducing the level of output by 23 percent.³¹ Thus, municipal regulation failed to hold prices down. But the utilities wanted an even more rapid increase in their prices, so they successfully lobbied for state regulation under the theory that state regulators would be less pressured by local customer groups, than mayors and city councils would be.

These research results are consistent with Horace Gray’s earlier interpretation of public utility rate regulation as an anti-consumer, monopolistic, price-fixing scheme.

The Problem of “Excessive Duplication”

In addition to the economies of scale canard, another reason that has been given for granting monopoly franchises to “natural monopolies” is that allowing too many competitors is too disruptive. It is too costly to a community, the argument goes, to allow several different water suppliers, electric power producers, or cable TV operators to dig up the streets. But as Harold Demsetz has observed:

[T]he problem of excessive duplication of distribution systems is attributable to the failure of communities to set a proper price on the use of these scarce resources. The right to use publicly owned thoroughfares is the right to use a scarce resource. The absence of a price for the use of these resources, a price high enough to reflect the opportunity costs of such alternative uses as the servicing of uninterrupted traffic and unmarred views, will lead to their overutilization. The setting of an appropriate fee for the use of these resources would reduce the degree of duplication to optimal levels.³²

³⁰George Stigler and Claire Friedland, “What Can Regulators Regulate? The Case of Electricity,” *Journal of Law and Economics* (October 1962): 1–16.

³¹Gregg A. Jarrell, “The Demand for State Regulation of the Electric Utility Industry,” *Journal of Law and Economics* (October 1978): 269–95.

³²Demsetz, *Efficiency, Competition, and Policy*, p. 81.

Thus, just as the problem with “natural” monopolies is actually caused by government intervention, so is the “duplication of facilities” problem. It is created by the failure of governments to put a price on scarce urban resources. More precisely, the problem is really caused by the fact that governments own the streets under which utility lines are placed, and that the impossibility of rational economic calculation within socialistic institutions precludes them from pricing these resources appropriately, as they would under a private-property competitive-market regime. Contrary to Demsetz’s claim, rational economic pricing in this case is impossible precisely because of government ownership of roads and streets. Benevolent and enlightened politicians, even ones who have studied at the feet of Harold Demsetz, would have no rational way of determining what prices to charge.

Murray Rothbard explained all this more than 25 years ago:

The fact that the government must give permission for the use of its streets has been cited to justify stringent government regulations of ‘public utilities,’ many of which (like water or electric companies) must make use of the streets. The regulations are then treated as a voluntary *quid pro quo*. But to do so overlooks the fact that governmental ownership of the streets is itself a permanent act of intervention. Regulation of public utilities or of any other industry discourages investment in these industries, thereby depriving consumers of the best satisfaction of their wants. For it distorts the resource allocations of the free market.³³

The so-called “limited-space monopoly” argument for franchise monopolies, Rothbard further argued, is a red herring, for how many firms will be profitable in any line of production “is an institutional question and depends on such concrete data as the degree of consumer demand, the type of product sold, the physical productivity of the processes, the supply and pricing of factors, the forecasting of entrepreneurs, etc. Spatial limitations may be unimportant.”³⁴

In fact, even if spatial limitations do allow only one firm to operate in a particular geographical market, that does not necessitate monopoly, for “monopoly” is “a meaningless appellation, unless monopoly price is achieved,” and “All prices on a free market are competitive.”³⁵ Only government intervention can generate monopolistic prices.

³³Murray N. Rothbard, *Power and Market: Government and the Economy* (Kansas City: Sheed Andrews and McMeel, 1977), pp. 75–76.

³⁴Murray N. Rothbard, *Man, Economy, and State: A Treatise on Economic Principles* (Auburn, Ala.: Ludwig von Mises Institute, 1993), p. 619.

³⁵*Ibid.*, p. 620.

The only way to achieve a free-market price that reflects true opportunity costs and leads to optimal levels of “duplication” is through free exchange in a genuinely free market, a sheer impossibility without private property and free markets.³⁶ Political fiat is simply not a feasible substitute for the prices that are determined by the free market because rational economic calculation is impossible without markets.

Under private ownership of streets and sidewalks, individual owners are offered a tradeoff of lower utility prices for the temporary inconvenience of having a utility company run a trench through their property. If “duplication” occurs under such a system, it is because freely-choosing individuals value the extra service or lower prices or both more highly than the cost imposed on them by the inconvenience of a temporary construction project on their property. Free markets necessitate neither monopoly nor “excessive duplication” in any economically meaningful sense.

Competition for the Field

The existence of economies of scale in water, gas, electricity, or other “public utilities” in no way necessitates either monopoly or monopoly pricing. As Edwin Chadwick wrote in 1859, a system of competitive bidding for the services of private utility franchises can eliminate monopoly pricing as long as there is competition “for the field.”³⁷ As long as there is vigorous bidding for the franchise, the results can be both avoidance of duplication of facilities and competitive pricing of the product or service. That is, bidding for the franchise can take place in the form of awarding the franchise to the utility that offers consumers the lowest price for some constant-quality of service (as opposed to the highest price for the franchise).

Harold Demsetz revived interest in the concept of “competition for the field” in a 1968 article.³⁸ The theory of natural monopoly, Demsetz pointed out, fails to “reveal the logical steps that carry it from scale economies in production to monopoly price in the market place.”³⁹ If one bidder can do the job at less cost than two or more, “then the bidder with the lowest bid price for the entire job will be awarded the contract, whether the good be cement, electricity, stamp vending machines, or whatever,

³⁶Ibid., p. 548.

³⁷Edwin Chadwick, “Results of Different Principles of Legislation and Administration in Europe of Competition for the Field as Compared With Competition Within the Field of Service,” *Journal of the Statistical Society of London* 22 (1859): 381–420.

³⁸Harold Demsetz, “Why Regulate Utilities?” *Journal of Law and Economics* (April 1968): 55–65.

³⁹Ibid.

but the lowest bid price need not be a monopoly price. . . . The natural monopoly theory provides no logical basis for monopoly prices.”⁴⁰

There is no reason to believe that the bidding process will not be competitive. Hanke and Walters have shown that such a franchise bidding process operates very efficiently in the French water supply industry.⁴¹

The Natural Monopoly Myth: Electric Utilities

According to natural monopoly theory, competition cannot persist in the electric utility industry. But the theory is contradicted by the fact that competition has in fact persisted for decades in dozens of U.S. cities. Economist Walter J. Primeaux has studied electric utility competition for more than 20 years. In his 1986 book, *Direct Utility Competition: The Natural Monopoly Myth*, he concludes that in those cities where there is direct competition in the electric utility industries:

- Direct rivalry between two competing firms has existed for very long periods of time—for over 80 years in some cities;
- The rival electric utilities compete vigorously through prices and services;
- Customers have gained substantial benefits from the competition, compared to cities where there are electric utility monopolies;
- Contrary to natural monopoly theory, costs are actually lower where there are two firms operating;
- Contrary to natural monopoly theory, there is no more excess capacity under competition than under monopoly in the electric utility industry;
- The theory of natural monopoly fails on every count: competition exists, price wars are not “serious,” there is better consumer service and lower prices with competition, competition persists for very long periods of time, and consumers themselves prefer competition to regulated monopoly; and
- Any consumer satisfaction problems caused by dual power lines are considered by consumers to be less significant than the benefits from competition.⁴²

Primeaux also found that although electric utility executives generally recognized the consumer benefits of competition, they personally preferred monopoly!

⁴⁰Ibid.

⁴¹Steve Hanke and Stephen J. K. Walters, “Privatization and Natural Monopoly: The Case of Waterworks,” *The Privatization Review* (Spring 1987): 24–31.

⁴²Walter J. Primeaux, Jr., *Direct Electric Utility Competition: The Natural Monopoly Myth* (New York: Praeger, 1986), p. 175.

Ten years after the publication of Primeaux's book, at least one state—California—is transforming its electric utility industry “from a monopoly controlled by a handful of publicly held utilities to an open market.”⁴³ Other states are moving in the same direction, finally abandoning the baseless theory of natural monopoly in favor of natural competition.⁴⁴

- The Ormet Corporation, an aluminum smelter in West Virginia, obtained state permission to solicit competitive bids from 40 electric utilities;
- Alcan Aluminum Corp. in Oswego, New York has taken advantage of technological breakthroughs that allowed it to build a new power generating plant next to its mill, cutting its power costs by two thirds. Niagara Mohawk, its previous (and higher priced) power supplier, is suing the state to prohibit Alcan from using its own power;
- Arizona political authorities allowed Cargill, Inc. to buy power from anywhere in the West; the company expects to save \$8 million per year;
- New federal laws permit utilities to import lower-priced power, using the power lines of other companies to transport it;
- Wisconsin Public Service commissioner Scott Neitzel recently declared, “free markets are the best mechanism for delivering to the consumer . . . the best service at the lowest cost”;
- The prospect of future competition is already forcing some electric utility monopolies to cut their costs and prices. When the TVA was faced with competition from Duke Power in 1988, it managed to hold its rates steady without an increase for the next several years.

The potential benefits to the U.S. economy from demonopolization of the electric utility industry are enormous. Competition will *initially* save consumers at least \$40 billion per year, according to utility economist Robert Michaels.⁴⁵ It will also spawn the development of new technologies that will be economical to develop because of lower energy costs. For example, “automakers and other metal benders would make much more intensive use of laser cutting tools and laser welding machines, both of which are electron guzzlers.”⁴⁶

⁴³“California Eyes Open Electricity Market,” *The Washington Times*, May 27, 1995, p. 2.

⁴⁴The following information is from Toni Mack, “Power to the People,” *Forbes*, June 5, 1995, pp. 119–26.

⁴⁵*Ibid.*, p. 120.

⁴⁶*Ibid.*, p. 126.

The Natural Monopoly Myth: Cable TV

Cable television is also a franchise monopoly in most cities because of the theory of natural monopoly. But the monopoly in this industry is anything but "natural." Like electricity, there are dozens of cities in the U.S. where there are competing cable firms. "Direct competition . . . currently occurs in at least three dozen jurisdictions nationally."⁴⁷ The existence of long-standing competition in the cable industry gives the lie to the notion that that industry is a "natural monopoly" and is therefore in need of franchise monopoly regulation. The cause of monopoly in cable TV is government regulation, not economies of scale. Although cable operators complain of "duplication," it is important to keep in mind that "while overbuilding an existing cable system can lower the profitability of the incumbent operator, it unambiguously improves the position of consumers who face prices determined not by historical costs, but by the interplay of supply and demand."⁴⁸

Also like the case of electric power, researchers have found that in those cities where there are competing cable companies prices are about 23 percent below those of monopolistic cable operators.⁴⁹ Cablevision of Central Florida, for example, reduced its basic prices from \$12.95 to \$6.50 per month in "duopoly" areas in order to compete. When Telestat entered Riviera Beach, Florida, it offered 26 channels of basic service for \$5.75, compared to Comcast's 12-channel offering for \$8.40 per month. Comcast responded by upgrading its service and dropping its prices.⁵⁰ In Presque Isle, Maine, when the city government invited competition, the incumbent firm quickly upgraded its service from only 12 to 54 channels.⁵¹

In 1987 the Pacific West Cable Company sued the city of Sacramento, California on First Amendment grounds for blocking its entry into the cable market. A jury found that "the Sacramento cable market was not a natural monopoly and that the claim of natural monopoly was a sham used by defendants as a pretext for granting a single cable television franchise . . . to promote the making of cash payments and provision of 'in-kind' services . . . and to obtain increased campaign contributions."⁵² The city was forced to adopt a competitive cable policy,

⁴⁷Thomas Hazlett, "Duopolistic Competition in Cable Television: Implications for Public Policy," *Yale Journal on Regulation* 7 (1990).

⁴⁸*Ibid.*

⁴⁹*Ibid.*

⁵⁰*Ibid.*

⁵¹Thomas Hazlett, "Private Contracting versus Public Regulation as a Solution to the Natural Monopoly Problem," in Robert W. Poole, ed., *Unnatural Monopolies: The Case for Deregulating Public Utilities* (Lexington, Mass.: Lexington Books, 1985), p. 104.

⁵²*Pacific West Cable Co. v. City of Sacramento*, 672 F. Supp. 1322 1349-40 (E.D. Cal. 1987), cited in Hazlett, "Duopolistic Competition."

the result of which was that the incumbent cable operator, Scripps Howard, dropped its monthly price from \$14.50 to \$10 to meet a competitor's price. The company also offered free installation and three months free service in every area where it had competition.

Still, the big majority of cable systems in the U.S. are franchise monopolies for precisely the reasons stated by the Sacramento jury: they are mercantilistic schemes whereby a monopoly is created to the benefit of cable companies, who share the loot with the politicians through campaign contributions, free air time on "community service programming," contributions to local foundations favored by the politicians, stock equity and consulting contracts to the politically well connected, and various gifts to the franchise authorities.

In some cities, politicians collect these indirect bribes for five to ten years or longer from multiple companies before finally granting a franchise. They then benefit from part of the monopoly rents earned by the monopoly franchisee. As former FCC chief economist Thomas Hazlett, who is perhaps the nation's foremost authority on the economics of the cable TV industry, has concluded, "we may characterize the franchising process as nakedly inefficient from a welfare perspective, although it does produce benefits for municipal franchisers."⁵³ The barrier to entry in the cable TV industry is not economies of scale, but the political price-fixing conspiracy that exists between local politicians and cable operators.

The Natural Monopoly Myth: Telephone Service

The biggest myth of all in this regard is the notion that telephone service is a natural monopoly. Economists have taught generations of students that telephone service is a "classic" example of market failure and that government regulation in the "public interest" was necessary. But as Adam D. Thierer recently proved, there is nothing at all "natural" about the telephone monopoly enjoyed by AT&T for so many decades; it was purely a creation of government intervention.⁵⁴

Once AT&T's initial patents expired in 1893, dozens of competitors sprung up. "By the end of 1894 over 80 new independent competitors had already grabbed 5 percent of total market share . . . after the turn of the century, over 3,000 competitors existed."⁵⁵ In some states there were over 200 telephone companies operating simultaneously. By 1907, AT&T's competitors had captured 51 percent of the telephone market and prices

⁵³Thomas Hazlett, "Duopolistic Competition in Cable Television."

⁵⁴Adam D. Thierer, "Unnatural Monopoly: Critical Moments in the Development of the Bell System Monopoly," *Cato Journal* (Fall 1994): 267-85.

⁵⁵*Ibid.*, p. 270.

were being driven sharply down by the competition. Moreover, there was no evidence of economies of scale, and entry barriers were obviously almost nonexistent, contrary to the standard account of the theory of natural monopoly as applied to the telephone industry.⁵⁶

The eventual creation of the telephone monopoly was the result of a conspiracy between AT&T and politicians who wanted to offer "universal telephone service" as a pork-barrel entitlement to their constituents. Politicians began denouncing competition as "duplicative," "destructive," and "wasteful," and various economists were paid to attend congressional hearings in which they somberly declared telephony a natural monopoly. "There is nothing to be gained by competition in the local telephone business," one congressional hearing concluded.⁵⁷

The crusade to *create* a monopolistic telephone industry by government fiat finally succeeded when the federal government used World War I as an excuse to nationalize the industry in 1918. AT&T still operated its phone system, but it was controlled by a government commission headed by the Postmaster General. Like so many other instances of government regulation, AT&T quickly "captured" the regulators and used the regulatory apparatus to eliminate its competitors. "By 1925 not only had virtually every state established strict rate regulation guidelines, but local telephone competition was either discouraged or explicitly prohibited within many of those jurisdictions."⁵⁸

The complete demise of competition in the industry, Thierer concludes, was brought about by the following forces: exclusionary licensing policies; protected monopolies for "dominant carriers"; guaranteed revenues or regulated phone companies; the mandated government policy of "universal telephone entitlement" which called for a single provider to more easily carry out regulatory commands; and rate regulation designed to achieve the socialistic objective of "universal service."

That free-market competition was the source of the telephone monopoly in the early twentieth century is the biggest lie ever told by the economics profession. The free market never "failed"; it was government that failed to permit free-market competition as it concocted its corporatist scheme to the benefit of the phone companies, at the expense of consumers and potential competitors.

⁵⁶Ibid.

⁵⁷G. H. Loeb, "The Communications Act Policy Toward Competition: A Failure to Communicate," *Duke Law Journal* 1 (1978): 14.

⁵⁸Thierer, "Unnatural Monopoly: Critical Moments in the Development of the Bell System Monopoly," p. 277.

Conclusions

The theory of natural monopoly is an economic fiction. No such thing as a “natural” monopoly has ever existed. The history of the so-called public utility concept is that the late-nineteenth- and early-twentieth-century “utilities” competed vigorously and, like all other industries, they did not like competition. They first secured government-sanctioned monopolies, and *then*, with the help of a few influential economists, constructed an *ex post* rationalization for their monopoly power.

This has to be one of the greatest corporate public relations coups of all time. “By a soothing process of rationalization,” wrote Horace M. Gray more than 50 years ago, “men are able to oppose monopolies in general but to approve certain types of monopolies . . . Since these monopolies were ‘natural’ and since nature is beneficent, it followed that they were ‘good’ monopolies . . . Government was therefore justified in establishing ‘good’ monopolies.”⁵⁹

In industry after industry, the natural monopoly concept is finally eroding. Electric power, cable TV, telephone services, and the mail, are all on the verge of being deregulated, either legislatively or *de facto*, due to technological change. Introduced in the U.S. at about the same time communism was introduced to the former Soviet Union, franchise monopolies are about to become just as defunct. Like all monopolists, they will use every last resource to lobby to maintain their monopolistic privileges, but the potential gains to consumers of free markets are too great to justify them. The theory of natural monopoly is a nineteenth-century economic fiction that defends nineteenth-century (or eighteenth century, in the case of the U.S. Postal Service) monopolistic privileges, and has no useful place in the twenty-first-century American economy.

⁵⁹Gray, “The Passing of the Public Utility Concept,” p. 10.

New Light on the Prehistory of the Theory of Banking and the School of Salamanca

Jesús Huerta de Soto

As is known, Murray N. Rothbard was one of the theorists who defended with the most creativity and coherence the need for free banking subject to general legal principles, in other words, banking with a cash ratio of 100 percent of demand deposits. Likewise, he was one of the first theorists to stress the great influence which the theoretical contributions of the Spanish scholastics of the University of Salamanca in the sixteenth and seventeenth centuries were to have as the direct predecessors of the Austrian School of Economics.¹

We feel that perhaps one of the greatest tributes which can be paid to Murray N. Rothbard is to show how the theorists of the School of Salamanca, whose intellectual activity took place from the reign of Carlos V in the sixteenth century onwards, developed an incipient theory on the legitimate practice of banking which coincides, to a great extent, with the contributions on this subject by the Austrian School in general and, particularly, by Murray N. Rothbard.

The analysis of banking during the reign of Carlos V is paradigmatic for several reasons. Firstly, because the massive inflow of precious metals from America caused the economic center of gravity to move, at least temporarily, from the mercantile cities in the north of Italy towards Spain, specifically Seville, and the other Spanish

*Jesús Huerta de Soto is professor titular de economía política, Universidad Complutense de Madrid, Spain.

¹See Murray N. Rothbard, "New Light on the Prehistory of the Austrian School," Edwin G. Dolan, ed., *The Foundations of Modern Austrian Economics* (Kansas City: Sheed and Ward, 1976), pp. 52–74. I have a letter from F. A. Hayek dated January 7, 1979, in which he states that, apart from Raymond de Roover, the researchers to whom we owe the establishment of the link between the School of Salamanca and the Austrian School are, chronologically, H.M. Robertson, *Aspects of the Rise of Economic Individualism* (Cambridge: Cambridge at the University Press, 1933); Marjorie Grice-Hutchinson, *The School of Salamanca* (Oxford: Clarendon Press, 1952); and, especially, Murray N. Rothbard in his above mentioned article.

The Review of Austrian Economics Vol. 9, No. 2 (1996): 59–81
ISSN 0889-3047

markets. Secondly, because Carlos V's constant need for cash, which was the result of his extravagant imperial policy, led him to continually finance his exploits through the banking system, taking advantage of the liquidity which it provided him without any scruples. The traditional complicity between bankers and governors which, although there was some degree of dissimulation, had already become a general rule was thus taken to its utmost limit by Carlos V. Moreover, he could not avoid the bankruptcy of the royal treasury which, logically, had pernicious consequences for the Spanish economy in general and, in particular, for the bankers who had financed him. All these events led the sharpest minds of the era, the theorists of the School of Salamanca, to begin to reflect on financial and banking activities and, as a result, we have a series of enormously valuable theoretical analyses which should be studied in detail. We analyze each of these aspects below.

The Development of the Banks in Seville

Thanks to the work of Ramón Carande,² we know about the development of private banking in Seville during the reign of Carlos V in some detail. Carande explains that he was able to carry out his research as a result of the list of bankers drawn up in relation to the confiscation of precious metals by the *Casa de Contratación* (Trade House) of Seville in 1545. The unsatisfactory situation of the treasury meant that Carlos V, violating the most elementary legal principles, resorted to appropriating money from where it was most readily available: deposits in the safes of the bankers of Seville. It is true that these bankers, as we will see later, also violated legal principles in relation to the demand-deposit contract (i.e., deposit of fungible money) and used a large part of the deposits received for their own business. However, it is no less true that the inauspicious imperial policy, by transgressing the most elementary principles of property rights and directly confiscating the stocks of money kept in the vaults, merely provided an even bigger incentive for the bankers to invest the greater part of the deposits received in loans, which became a habitual practice: if, in the final analysis, there was no guarantee that the public authorities would respect the part of the cash reserve which was kept in the bank (and experience showed that, when times were difficult, the Emperor did not hesitate to confiscate this reserve and substitute it by compulsory loans to the Crown), it was preferable to devote the greater part of the deposits to loans to private industry and commerce, thus avoiding expropriation and obtaining a greater profitability.

²Ramón Carande, *Carlos V y sus Banqueros*, 3 vols. (Barcelona and Madrid: Editorial Crítica, 1987).

In any case, this policy of confiscation is perhaps the most extreme manifestation of the public authorities' traditional policy of taking advantage of illegitimate banking profits by expropriating the assets of those who, by legal obligation, should best guard and preserve the deposits of third parties. It is understandable, therefore, that the governors, as the main beneficiaries of the bankers illegitimate activity, ended up justifying it and granting it all kinds of privileges so that it could continue to act, with a fractional-cash ratio, outside the framework of general legal principles.

In his *magnum opus*, *Carlos V y sus Banqueros* ("Carlos V and his Bankers"), Ramón Carande lists the most important bankers in the Seville of Carlos V, specifically the Espinosas, Domingo de Lizarrazas and Pedro de Morga, together with other, less important, bankers, such as Cristóbal Francisquín, Diego Martínez, Juan Íñiguez and Octavio de Negrón. All of them inexorably became bankrupt, basically due to having insufficient liquidity to meet the withdrawal of demand deposits which had been placed with them. This shows that they worked with a fractional-cash ratio, thanks to the license or privilege which they had obtained from the Municipality of Seville and from Carlos V himself. We have no information on the percentage of reserves they held, but we do know that, on many occasions, they invested in their own shipping businesses (fleets which traded with America) and the collection of taxes which were a tremendous temptation because, if the conditions were favorable, the profits were very significant. Moreover, the successive confiscations of precious metals deposited with the bankers merely provided a greater incentive for their illegitimate behavior. Thus, the Espinosas went bankrupt in 1579 and the main partners were imprisoned. The bankruptcy of Domingo de Lizarrazas occurred on March 11, 1553, when he could not meet a payment of more than six and a half million maravedis. Pedro de Morga, who began operations in 1553, went bankrupt in 1575, during the second bankruptcy of King Felipe II. The rest of the less important bankers met the same fate and, in this respect, it is interesting to note the presence and comment of Thomas Gresham, who travelled to Seville with instructions to withdraw three hundred and twenty thousand ducats in cash, for which he had obtained the necessary license from the Emperor and from Queen María. Gresham was astounded to find that, in the city which received the treasures of the Indies, money was very scarce, as it was in the trade markets, and he feared that, upon withdrawal of the funds to which the orders he was bearing referred, all the banks of the city would suspend payments.³ It is regrettable that Ramón

³Finally, with a great deal of effort, he managed to obtain two hundred thousand ducats but, as he wrote, "I am afraid of causing the bankruptcy of all the banks of Seville."

Carande's theoretical analysis leaves so much to be desired and that his study interpreting the bankruptcy of these banks is based solely on anecdotal "explanations," such as "greed" for metals, which constantly placed the solvency of the bankers in a situation of crisis; the fact that the bankers carried out risky personal business deals, which continually implied heavy obligations (the chartering of vessels, overseas maritime trade, insurance transactions, various speculative types of business, etc.); and the royal treasury's repeated confiscations and need for liquidity. Nowhere is the true cause of the phenomenon mentioned: the inevitable recession and economic crisis resulting from the artificial boom caused by the inflation of precious metals from America and the artificial credit expansion without an adequate base of real saving, derived from the practice of banking with a fractional-cash ratio.

Fortunately, Carlo M. Cipolla has covered this gap in theory of Ramón Carande, at least partially, and has made a study interpreting the banking and economic crisis of the second half of the sixteenth century which, although it refers strictly to the Italian banks, is also directly applicable to the Spanish financial system, as the trading and financial circuits and flows between the two nations were, at that time, intimately related.⁴ Cipolla explains that the monetary supply (what is today called M1 or M2) in the second half of the sixteenth century included a large amount of "bank money" or deposits created out of nowhere by the bankers who did not keep a 100-percent-reserve ratio of the cash which had been deposited with them at demand by their clients. This led to a tremendous artificial thriving of the economy, which inevitably reversed in the second half of the sixteenth century when depositors began to experience economic difficulties and the wave of bankruptcies felt by the most important bankers in Florence began.

This expansionary phase was started in Italy, according to Cipolla, by the managers of the Ricci Bank, who used a significant part of the bank's newly created deposits to purchase public funds and grant credits. This policy of credit expansion dragged the other private banks along with it, if they wanted to be competitive and maintain their profits and market share. A state of credit euphoria was thus created, which gave rise to a great artificial expansion that soon began to revert. Thus, we

See Ramón Carande, *Carlos V y sus Banqueros*, vol. 1, pp. 299–23, and esp. pp. 315–16, which deals with Gresham's visit to Seville.

⁴See Carlo M. Cipolla's important article "La moneta en Florencia en el Siglo XVI," published in *El Gobierno y la Moneda: Ensayos de Historia Monetaria* (Barcelona: Editorial Critica, 1994), pp. 11–142; esp. pp. 96 onwards. This book is the Spanish edition of the work originally published in Italian with the title *Il Governo della Moneta: La Moneta a Firenze nel Cinquecento* (Bologna: Società editrice Il Mulino, 1990).

can read an edict of 1574 in which accusations are made against bankers who refuse to return deposits in cash and which proclaims the fact that they only “paid with ink.” They had increasing difficulties in returning deposits in ready money and a significant money shortage began to be perceived in the Venetian cities. The artisans could not withdraw their money or pay their debts and there was a heavy credit contraction (in other words, a deflation) and a deep economic crisis, which Cipolla analyzes in detail in his brilliant book. Cipolla’s analysis is, therefore, much more solid from a theoretical point of view than that of Ramón Carande, although it cannot be considered perfect, as it places the emphasis more on the crisis and the period of credit contraction than on the preceding phase of artificial credit expansion, which was the true origin of the evils and of which, in turn, the most intimate cause of the bankers’ violation of the obligation to guard and conserve intact 100 percent of the *tantundem* or equivalent of the original deposits.⁵

The School of Salamanca and the Banking Business: The Initial Contribution of Doctor Saravia de la Calle

The financial and banking phenomena that we are discussing made an impression on the outstanding minds of the members of the School of Salamanca who, according to the most reliable research, are the forerunners of the subjectivist conception developed by the Austrian School of Economics.⁶

⁵Cipolla tells us how the Ricci Bank, from the seventies onwards, was not able to meet the demand for payments in cash and, *de facto*, suspended payments, as it paid simply “with ink” or “with bank policies.” The authorities of Florence, looking only at the symptoms and trying, with typical good intentions, to resolve this worrying situation merely by decrees, imposed on bankers the obligation to pay their creditors in cash without any delay, but did not attack the fundamental causes of the phenomenon (the undue appropriation of the deposits as loans and failure to hold a 100-percent-cash ratio). This meant that the successive decrees issued met with inevitable failure and the crisis became gradually more serious until it broke out with its full virulence in the mid-1570s. *Ibid.*, pp. 102–3.

⁶Among others, the following have recently studied the contribution of the Spanish scholastics to economic theory: Lucas Beltrán in “Sobre los orígenes hispanos de la economía de mercado,” *Cuadernos del Pensamiento Liberal* 10, no.1 (1989): 5–38; Marjorie Grice-Hutchinson, *Economic Thought in Spain: Selected Essays of Marjorie Grice-Hutchinson*, Laurence S. Moss and Christopher K. Ryan, eds. (Aldershot, England: Edward Elgar, 1993); Jesús Huerta de Soto, “Génesis, esencia y evolución de la Escuela Austriaca de Economía,” *Estudios de Economía Política* (Madrid: Unión Editorial, 1994), pp. 17–55; and esp. and most recently Murray N. Rothbard, *Economic Thought before Adam Smith: An Austrian Perspective on the History of Economic Thought*, vol. 1 (Aldershot, England: Edward Elgar, 1995), pp. 101–27. The intellectual influence of the Spanish theorists of the School of Salamanca on the Austrian School is not, however, a pure coincidence or a mere whim of history. It originates from and exists because of the intimate historical, political and cultural relations which, as from the reigns of Carlos V and his brother Fernando I, arose between Spain and Austria and which were to continue for several centuries. In addition, Italy also played an important role in these relations, acting as an authentic cultural, economic and financial bridge over which the relations between

Chronologically, the first work we should mention, which is also, perhaps, the most relevant to our purpose, is *Instrucción de Mercaderes* ("Instruction of Merchants") by Doctor Saravia de la Calle, which was published in Medina del Campo in 1544.⁷ Saravia de la Calle is extraordinarily hard on the bankers, whom he describes as "hungry gluttons, who swallow everything, destroy everything, confuse everything, steal and dirty everything, like the harpies of Pineo."⁸ He tells us how the bankers "come out into the square and road with their table and chair and cashbox and book, like the whores to the brothel with their chair" and, having obtained the corresponding license and guarantee ordered by the laws of the kingdom, they devote themselves to obtaining deposits from the clients, to whom they offer bookkeeping and cashier services, paying into their account and even paying interest on such deposits.

With sound legal criteria, Saravia de la Calle says that receiving interest is incompatible with the nature of a demand deposit and that, in any case, a fee should be paid to the banker for keeping or guarding the money under his custody. He even reprehends harshly the clients of the banks who agree to enter into such deals with the bankers. He states: "And if you say, merchant, that you do not lend it, but that you place it (or deposit it), that is a greater mockery; who ever saw the depositary pay? He is usually paid for the safekeeping and the work of the deposit; much more than that, if you now place your money with the moneylender as a loan or as a deposit, in the same way as you take your part of the profit that the moneylender takes, you also take part of the blame, and even the greatest part."⁹ In Chapter 12 of his book, Saravia de la Calle also correctly distinguishes between the two radically different transactions which the banks carry out. On the one hand, the demand deposits, which the clients give, without any interest, to the bankers "to have them safer and to have them more at hand in order to deliver them to whom they are owed, and to free themselves from the burden and the work of reckoning and safekeeping, and also because, as thanks for this good deed which they do to the moneylenders in giving them their money, if it occurs that they have no money in the hands of the moneylender, the

the two furthest points of the Empire (Spain and Vienna) flowed. In this respect, Jean Berenguer's interesting book *El Imperio de los Habsburgos* (Barcelona: Editorial Critica, 1993), should be consulted, particularly pp. 133–35. This book is the Spanish edition of the French original which was entitled *Histoire de L'empire des Habsbourg 1273–1918* (Paris: Librairie Arthème, Fayard, 1990).

⁷Luis Saravia de la Calle, *Instrucción de Mercaderes* (Medina del Campo: Pedro de Castro, 1544); republished in *Colección de Joyas Bibliográficas* (Madrid, 1949).

⁸Ibid., p. 180.

⁹Ibid., p. 181.

moneylender accepts some overdrafts from them also without interest.”¹⁰ On the other hand, very different from these contracts are the term “deposits,” which are true loans and are characterized because they are given over a time period in exchange for interest. Saravia de la Calle, following the traditional canon law doctrine, deeply condemns this practice. Moreover, he clearly indicates that, in the case of the first type of demand-deposit contracts, the clients should pay the banker “because if they place monies in deposit, they must give for their safekeeping, not receive the profits given to them when they deposit monies or goods which must be safely kept.”¹¹ Saravia de la Calle therefore criticizes those clients who selfishly try to take advantage of the illegitimate activity of the bankers, entrusting them with their money in demand deposits and trying to obtain interest on it. He adds the following illustrative words: “he is not free from sin, at least venial sin, because he entrusted the deposit of his money to whom he knows will not keep his deposit, but will spend his money, like he who entrusts the maiden to the lecher or the delicacy to the glutton.”¹² Neither may the depositor clear his conscience by thinking that the banker will lend or use the money of others, but not his own as if “it is believed of him that he will probably keep that money of the deposit and will not lend it; and this probability cannot be thought of any of these moneylenders, but the contrary, that he will immediately lend it and deal and obtain earnings with it, because how can those who give 7 and 10 percent to those who give them monies leave the monies which are thus placed with them in deposit idle? And even if it were very certain that you do not sin (which it is not, but the contrary), it is very sure that the moneylender sins lending your monies, and that he steals the patrimony of your neighbors with your money.”¹³ The doctrine of Saravia de la Calle is, therefore, quite clear: the use of money that is deposited at demand with the bankers, which in their own interest grant loans, is illegitimate and implies a grave sin. This doctrine fully coincides with the doctrine which was established by the classical authors on Roman Law and arises naturally from the essence, purpose and legal nature of the contract for the demand deposit of money.¹⁴

Saravia de la Calle is also very emphatic when talking about the enormous profits that the bankers receive from their illegitimate behavior of appropriating the deposits of their depositors, instead of being

¹⁰Ibid., p. 195.

¹¹Ibid., p. 196.

¹²Ibid., p. 197.

¹³Ibid.

¹⁴See Jesús Huerta de Soto, “A Critical Analysis of Central Banks and Fractional-Reserve Free Banking from the Austrian School Perspective,” *Review of Austrian Economics* 8, no. 2 (1995): 25–38, esp. pp. 29–30 n. 6.

content with the lesser remuneration which they would receive for the simple safekeeping or custodial oversight of the deposits, like good fathers of families. Let us see how vividly he expresses this: "And if you receive wages, they should be moderate, for you to support yourselves, and not such excessive robberies that you build superb houses and purchase rich property, have excessive costs of family and servants, and hold great banquets and dress in such a costly fashion, especially when you were poor when you began to lend and left poor trades."¹⁵ Saravia de la Calle indicates how the bankers have a great tendency to become bankrupt, even making a brief theoretical analysis which shows that, after the expansionary phase resulting from the artificial expansion of the credits which these *logreros* ("moneylenders") grant, there inevitably comes a phase of recession in which the bad debts cause a chain of bankruptcies among the banks. And he adds that, "if the merchant does not pay the moneylender, it makes him bankrupt, and thus he suspends payments and all is lost, of all which, as is notorious, these moneylenders are the beginning, occasion and cause, *because if they did not exist, each person would trade with his money as he could and no more, and thus things would cost their just price and more than the price in cash would not be charged*. And, therefore, it would be a great advantage if the princes did not consent to them in Spain, as no other nation in the world consents to them, and banished them from their court and kingdom."¹⁶ As we know, it is not true that the governors of other nations had had more success in controlling the activity of the bankers or lenders than in Spain, as more or less the same was occurring everywhere and the kings ended up by granting privileges for the bankers to carry on their activities using the money of their depositors for their own interest, in return for also being able to take full or partial advantage of a banking system from which they obtained funds much more easily and rapidly than from taxes.

As a conclusion to his analysis, Saravia de la Calle states that "under no circumstance should the Christian give his monies to these lenders because, if he sins in giving them, as he always sins, he should cease it because it is his own sin; and if he does not sin, he should cease it in order to avoid the sin of the moneylender." Moreover, Saravia de la Calle adds that, if the bankers are not used, there will be the additional advantage that the depositors "will not be startled if the lender suspends payments; if he becomes bankrupt, as we see so commonly and our Lord God permits, they and their owners will be lost like an ill-gained thing."¹⁷ As we can see, Saravia de la Calle's analysis, in

¹⁵Saravia de la Calle, *Instrucción de Mercaderes*, p. 186.

¹⁶*Ibid.*, p. 190; emphasis added.

¹⁷*Ibid.*, p. 198.

addition to its ingenuity and humor, is impeccable and has no contradictions, except, perhaps, that it puts too much emphasis on the criticism of the bankers for charging interest, which violated the canonical prohibition of usury, rather than for their undue appropriation of the demand deposits placed with them by their clients.

Martín de Azpilcueta Navarro

Another writer who made a correct analysis of the contract for the demand deposit of money is Martín de Azpilcueta, better known as “Doctor Navarro,” in his book *Comentario Resolutorio de Cambios* (*Resolatory Commentary on Exchanges*), first published in Salamanca at the end of 1556. Martín de Azpilcueta expressly refers to “banking for safekeeping,” which is the monetary demand-deposit transaction performed by the banks. For Martín de Azpilcueta, banking for safekeeping or the demand-deposit contract is completely fair and consists of the banker being the “warden, depositary and guarantor of the monies, which those who give him or send to him give to him or bank with him for what may be necessary; and that he is obliged to pay to the merchants, or to the persons whom the depositors wish in such or such a way, for which he may licitly take his fair wage, either from the republic or from the depositors; because this trade and duty is useful to the republic and does not contain any iniquity, as it is fair that he who works earns his wage. And the banker works in receiving, holding in deposit and ready the money of so many merchants, and in writing, giving and keeping accounts with all of them, with great difficulty, and sometimes danger of errors in reckoning and other things. The same could be done with a contract with which a person committed himself to the others to receive and hold their money in deposit, give, pay and keep the accounts of all of them, as they tell him, etc., because this contract is of hire to another and from another of his works and employment, which is a designated, just and holy contract.”¹⁸ As we can see, for Martín de Azpilcueta, the contract of a demand deposit is a fully legitimate contract, which consists of entrusting the safekeeping, custody or deposit of money to a professional, the banker, who should care for it like the good father of a family, always keeping it at the depositor’s disposal and performing the cashier services requested by the latter on his behalf. In return, he will have the right to receive the appropriate remuneration of

¹⁸Martín de Azpilcueta, *Comentario Resolutorio de Cambios* (Madrid: Consejo Superior de Investigaciones Científicas, 1965), pp. 57–58. When studying the position of Doctor Navarro, I have worked with the first Spanish edition published by Andrés de Portonarijs in Salamanca in 1556, and also with the first Portuguese edition published by Ioam de Barreyra in Coimbra in 1560 with the title *Comentario Resolutorio de Onzenas*. The quotes contained in the main text appear in the Portuguese edition on pp. 77–80, and on pp. 74–75 of the first Spanish edition.

his services from the depositors. In effect, for Martín de Azpilcueta, *the depositors should pay the depositary or banker*, never *vice versa*, so that the depositors “pay the former for the work and care of the banker in receiving and keeping their money and doing his work” and, therefore, the bankers should perform “their trade cleanly and be content with a just wage, receiving it from those who owe it to them and whose money they safeguard and accounts they keep and not from those who do not owe it to them.”¹⁹ Moreover, in order to avoid confusion and make things perfectly clear, Martín de Azpilcueta, along the same lines as Doctor Saravia de la Calle which we have seen above, expressly condemns the clients who do not wish to pay for the services of custody of their deposits or even try to receive interest on them. Thus, Doctor Navarro concludes that “in this type of exchange, not only the bankers sin, but, even . . . those who give them monies for them to keep, and do the same. And afterwards they do not want to pay them anything, saying that that which they earn with their money, and will receive from those to whom they pay in cash, suffices for their wages. And if the bankers ask them for anything, they leave them and go to deal with others, and so that they do not leave them, the bankers renounce the wage owed to them and take it from who does not owe it.”²⁰

The Contribution of Tomás de Mercado

Tomás de Mercado, in his *Suma de Tratos y Contratos* (“Compilation of Deals and Contracts”), Seville, 1571, makes an analysis of the banking business which follows a very similar line to the two above authors. Firstly, he points out, following the correct doctrine, that the depositors should pay the bankers for the work of keeping their monetary demand deposits, concluding that “for all of them it is a common and general rule to be able to take wages from those who place money in their bank, either a certain amount each year or a certain amount for each thousand, as they serve them and keep their patrimony.”²¹ However, Tomás de Mercado ironically mentions that the bankers of Seville are so “generous” that they do not make any charge for the custody of the deposits, using the following words: “those of this city, it is true, are so regal and noble that they do not ask for or take any wage.”

¹⁹Ibid., pp. 60–61.

²⁰Ibid., p. 61.

²¹I quote the edition by the Instituto de Estudios Fiscales published in Madrid in 1977, edited and introduction by Nicolás Sánchez Albornoz, vol. 2, chap. 16, p. 479. There is another edition, by Restituto Sierra Bravo, published by the Editora Nacional in 1975, which includes the quote given in the main text on p. 401. The original edition was published in Seville in 1571 “en casa de Hernando Díez Impresor de Libros, en la calle de la Sierpe.”

And Tomás de Mercado remarks that the bankers of Seville do not need to charge anything as, with the large amount of money they obtain from the deposits, they carry out their private businesses, which are very lucrative. We should stress the fact that, in our opinion, the analysis made by Tomás de Mercado in this respect refers simply to an observation of a fact, and does not imply any acceptance of its legitimacy, as several modern critics appear to suggest. To the contrary, following the purest, classical-Roman doctrine and the essence of the legal nature of the contract for the demand deposit of money, Tomás de Mercado is the scholastic writer who most clearly shows that the transfer of property which occurs in the monetary bank deposit does not imply a parallel transfer of availability and, therefore, for practical purposes, a *full* transfer of property does not take place. Let us see how clearly he expresses this: “[the bankers] must understand that the money is not theirs but belongs to someone else, and that is not all, when they have it serve them, it ceases to serve its owner.” Tomás de Mercado adds that the bankers should be subject to two basic principles, the first, “not to leave the bank so bare that they cannot then pay the drafts which come, because, if they make it impossible to pay them, spending and employing money in investments and speculative earnings or other deals, it is certain that they sin . . . The second: they should not enter dangerous businesses, because they sin, even if they result favorably, due to the danger of behaving wrongly and doing grave damage to those who trusted them.”²²

Although it is true that, with these recommendations, it seems as if Tomás de Mercado would admit the use of a certain fractional-reserve ratio, the fact is that he is very forthright in expressing his legal opinion that, in the final analysis, the money of the deposits does not belong to the bankers but to the depositors and when he says, moreover, that no banker heeds his two recommendations: “but in the case of earning, when it is comfortable, it is very difficult to restrain avarice, none of them heeds these two warnings, nor meets these conditions.”²³ Therefore, he considers very favorably the enactment of a rule prohibiting the bankers from doing private business, in order to remove the temptation implied by financing them indefinitely with the money obtained from demand deposits.

In addition, elsewhere in *Suma de Tratos y Contratos*, at the end of chapter 4, Tomás de Mercado tells us that the bankers of Seville act as depositaries for the money and precious metals of the merchants from

²²Tomás de Mercado, *Suma de Tratos y Contratos*, vol. 2, p. 480 of the edition of Instituto de Estudios Fiscales and p. 401 of the edition of Restituto Sierra Bravo.

²³*Ibid.*, p. 480.

the fleet of the Indies and that, with such substantial deposits “they make great investments” and obtain lucrative profits, now without expressly condemning this type of activity, although it is true that the passage in question is a description of a state of affairs rather than an analysis of the legitimacy of the situation. This analysis is made in much more depth in chapter 16, which we have discussed above. Tomás de Mercado concludes, moreover, that the bankers “also become involved in giving and taking in exchange and in collecting, and a banker in this republic covers a whole world and embraces more than the ocean, although sometimes he leaves so many loose ends that everything goes to ruin.”²⁴

The Cases of Domingo de Soto, Luis de Molina and Juan de Lugo

The scholastics who are most confused on the doctrinal treatment of the contract for the monetary bank deposit are Domingo de Soto and, above all, Luis de Molina and Juan de Lugo. In fact, these theorists allowed themselves to be influenced by the wrong medieval tradition of the glossators and, especially, by the doctrinal confusion which developed due to the concept of *depositum confessatum*. This was simply a loan which was disguised as a deposit in order to evade the canonical prohibition on charging interest, as this practice was considered acceptable if there was a (fictitious) delay on the part of the depositary. In fact, de Soto and, above all, Molina wrongly considered that the demand deposit was merely a “loan” which transferred not only the property, but also the full availability of the deposits to the banker and, therefore, it could be considered legitimate to use them as loans, provided that these were made “prudently.” It may be implied that Domingo de Soto was the first to uphold this thesis, although very indirectly. In fact, in Book VI, Question XI of his work on *De Iustitia et Iure* (*On Justice and Law*) (1556), we can read that, among the bankers, there was “the custom, it is said, that if a merchant makes a bank deposit in cash, as a result thereof the banker answers for a higher amount. I delivered ten thousand to the moneychanger, then he will answer for me for 12, perhaps 15; because it is very good earning for the banker to have money in cash. Neither is any vice found therein.”²⁵ Another case of typical credit creation which seems to be

²⁴This is the quote from Mercado that Ramón Carande includes in Vol. 1 of *Carlos V y sus Banqueros* in the introduction to his analysis of the bankers of Seville and the crisis which led them all into bankruptcy. See Tomás de Mercado, *Suma de Tratos y Contratos*, vol. 2, pp. 381–82 of the 1977 edition by Instituto de Estudios Fiscales and p. 321 of the edition by Restituto Sierra Bravo.

²⁵According to Restituto Sierra Bravo, *El Pensamiento Social y Económico de la Escolástica*, vol. 1 (Madrid: Consejo Superior de Investigaciones Científicas, 1975), p. 215, this phrase of Domingo de Soto implies his acceptance of the banking business with a fractional-reserve ratio.

admitted by Domingo de Soto is that of a loan in the form of the discount of bills financed against the demand deposits of the clients.

But perhaps the member of the School of Salamanca who upheld the most erroneous doctrine in relation to the contract for the demand deposit of money made by the bankers was Luis de Molina.²⁶ In fact, Luis de Molina, in his *Tratado sobre los Cambios* ("Treatise on Exchanges"), upholds the medieval doctrine that the demand deposit is merely a "loan" contract in favor of the banker, which transfers not only the property, but also the full availability of the thing and, therefore, the banker can legitimately use it in his own interest, in the form of loans or in any other way. Let us see how he expresses his argument: "because these bankers, like all the rest, are the true owners of the money which is deposited in their banks, in which they are greatly differentiated from the other depositaries . . . in such a way that they receive it as a loan with no rights attached and, therefore, at their own risk"; and, elsewhere, further on, he again repeats that "such a deposit is really a loan, as has been said, and the property of the money deposited passes to the banker and, therefore, in the event that it perishes, it perishes for the banker."²⁷ This doctrinal position is a clear lapse and contradicts what the writer himself says in his other work, *Tratado sobre los Préstamos y la Usura* ("Treatise on Loans and Usury"), where he warns that the *term* is an essential element of any loan contract and that, if the time for which a loan may be held is not expressly stated (as happens in a demand deposit) and no date has been fixed for its repayment, "it will be necessary to abide by what the judge judges as to the time for which it may be held."²⁸ In addition, Luis de Molina ignores the fact that the nature and legal essence of the demand deposit contract has nothing to do with the loan contract and, therefore, his doctrine which tries to identify the one with the other is a clear regression, not only in relation to the much more coherent positions of Saravia de la Calle and Azpilcueta Navarro, but also in relation to the true legal nature of the contract as it had been developed by the old Roman legal tradition. It is, therefore, surprising that so clear and profound a mind as that of Luis de Molina did

²⁶It is very significant that various authors, among them Marjorie Grice-Hutchinson, are in doubt as to whether Luis de Molina should be included among the theorists of the School of Salamanca: "The inclusion of Molina in the School seems to me now to be more dubious." Marjorie Grice-Hutchinson, "The Concept of the School of Salamanca: Its Origins and Development", Chapter 2 of *Economic Thought in Spain: Selected Essays of Marjorie Grice-Hutchinson*, p. 25.

²⁷Luis de Molina, *Tratado sobre los Cambios*, edition and introduction by Francisco Gómez Camacho (Madrid: Instituto de Estudios Fiscales, 1990), pp. 137–40.

²⁸Ibid., p. 13.

not realize how extremely dangerous it was to accept the violation of the general legal principles on the bank deposit and to say that "it never occurs that all the depositors need their money in such a way that they do not leave many thousands of ducats in deposit with which the bankers may do business to obtain profits or losses."²⁹ Molina did not realize that not only the objective or essential purpose of the contract, which is the safekeeping and custody of money, is thus violated, but that all kinds of illicit businesses and abuses are encouraged, which inexorably generate an economic recession and bankruptcy of the bankers. If the traditional legal principle which requires the continual safekeeping of 100 percent of the *tantundem* in favor of the depositor is not met, there is no clear guideline for avoiding bankruptcy of the bankers. And it is evident that such superficial and vague suggestions as to "try to act with prudence" or "not get involved in dangerous business" are insufficient to avoid the very prejudicial economic and social effects of fractional-reserve banking. However, Luis de Molina does at least take the trouble to point out that "a warning should be given that (the bankers) are in mortal sin if they commit the money they hold in deposit in their businesses to such an extent that they then find themselves unable to deliver, at the appropriate moment, the amounts which the depositors request or order to be paid against the money which they hold in deposit . . . Likewise, they are in mortal sin if they engage in such businesses that they are in danger of reaching a situation in which they cannot pay the deposits. For example, if they send so much merchandise overseas that, if the ship is wrecked, or if it is captured by pirates, it is not possible for them to pay the deposits, even if they sell their patrimony. *And not only are they in mortal sin when the business ends badly, but even if the outcome is favorable. And this is due to the danger to which they expose themselves of causing damage to the depositors and guarantors which they themselves contributed for the deposits.*"³⁰ We consider this warning by Luis de Molina to be commendable, but we also consider it extraordinary that, it seems, he did not realize that it is, in the final analysis, intimately contradictory to his express acceptance of fractional-reserve banking, provided that the bankers practice it with "prudence." And it does not matter how prudent the bankers are, the only way to avoid risks and to guarantee that the depositors' money is always at their disposal is by maintaining a cash ratio of 100 percent of all the demand deposits received.

²⁹Ibid., p. 137.

³⁰Ibid., pp. 138–39; emphasis added.

After Molina, the only author who upheld an analogous position on banking is Juan de Lugo,³¹ also a Jesuit. In our opinion, this can lead us to consider that, in relation to banking, there were two schools of thought within the School of Salamanca: one, well-founded and doctrinally correct (close to the future “currency school”), to which Saravia de la Calle, Azpilcueta Navarro and Tomás de Mercado belonged; and another, more inclined towards the capriciousness of the inflationist doctrine and the fractional reserve (close to the future “banking school”), represented by Luis de Molina, Juan de Lugo and, to a much less extent, Domingo de Soto. We will study these two points of view in more detail in the next section.

The School of Salamanca: Banking and Currency View

The contributions of the theorists of the School of Salamanca in the monetary field are important and have been the subject of detailed studies.³²

The first scholastic treatise which dealt with money was written by Diego de Covarrubias y Leyva and published in 1550 under the title *Veterum Collatio Numismatum* (“Compilation on old moneys”). In this work, the famous Bishop of Segovia studied the history of the devaluation of the Castilian maravedi and compiled a large number of statistics on the evolution of prices. Although the essential ideas of the quantity theory of money are already implicit in Covarrubias’ treatise, he does not put forward an explicitly articulated monetary theory.³³ Some years were to pass before, in 1556, Azpilcueta Navarro expressed, for the first time, clearly and convincingly that the increase in prices, if one prefers, the decrease in the purchasing power of money, was the result of the increase in the monetary supply which was taking place in Castile as a result of the massive inflow of precious metals from America.

³¹R.P. Joannis de Lugo Hispalensis, S.I., *Disputationum de Iustitia et Iure, Tomus Secundus* (Lugduni, 1642), Disp. 28, Sec. 5, pp. 406–7. I would like to thank the Jesuit Father Prof. Enrique M. Ureña and the Dominican Father Rodrigo T. Hidalgo who provided me with different copies of de Lugo’s original book.

³²See, above all, the doctoral thesis which Marjorie Grice-Hutchinson published under the direction of F.A. Hayek with the title of *The School of Salamanca: Readings in Spanish Monetary Theory, 1544–1605*; Murray N. Rothbard, “New Light on the Prehistory of the Austrian School,” pp. 52–74; Alejandro A. Chafuen, *Christians for Freedom: Late-Scholastic Economics* (San Francisco: Ignatius Press, 1986), pp. 74–86; and Murray N. Rothbard, *Economic Thought before Adam Smith: An Austrian Perspective on the History of Economic Thought*, vol. 1, pp. 101–27.

³³The edition which I have used is the *Omnia Opera*, published in Venice in 1604, which includes the treatise on money by Diego de Covarrubias in vol. 1, under the full title of *Veterum Collatio Numismatum, cum his, quae modo expendantur, publica, et Regia auctoritate perpensa*, pp. 669–710. This work by Diego de Covarrubias is often quoted by Davanzati and, at least once, in chap. 2 of Galiani’s famous *Della Moneta*, written in 1750. Also, and most significantly, it is quoted by Carl Menger, *Grundsätze der Volkswirtschaftslehre* (Vienna: Wilhelm Braumüller, 1871), p. 257; English edition, *Principles of Economics* (New York and London: New York University Press, 1981), p. 317.

In fact, the relationship between the amount of money in circulation and prices is impeccably expressed by Martín de Azpilcueta, for whom "in the lands where there is a great shortage of money, all the other things which may be sold, even the labor and work of men, are given for less money than in places where there is an abundance; as can be seen from experience, in France, where there is less money than in Spain, bread, wine, clothing, labor and work cost much less; and even in Spain, at the time when there was less money, the things which could be sold and the labor and work of men were given for much less than after the Indies were discovered and covered her with gold and silver. *The cause of which is that money is worth more where and when it is lacking than where and when it is in abundance.*"³⁴

However, in contrast to the deep and detailed studies which have been made of the monetary theory of the School of Salamanca, up to now there has been practically no effort to analyze the position of the scholastics in relation to banking.³⁵ And nevertheless, as we have seen in preceding sections, the theorists of the School of Salamanca made a very acute analysis of banking practices and, to a great extent, were the forerunners of the opposing positions which, more than two centuries later, were reproduced in England in the controversy between the members of the *banking school* and those of the *currency school*.

In fact, we have already set out the profoundly critical treatment of fractional-reserve banking which we owe, mainly, to Doctor Saravia de la Calle and which is included in the final chapters of his *Instrucción de Mercaderes*. Martín de Azpilcueta and Tomás de Mercado also developed a rigorous and very demanding critical analysis of banking activities which, although it did not reach the degree of criticism of Saravia de la Calle, included an impeccable treatment of the demands which, in accordance with justice, should be observed in the monetary-bank-deposit contract. For this reason, and due to their rigorous critical analysis of banking, we may consider this first group of authors (most of them Dominicans) to form part of an incipient *currency school*, which had been developed from the start within the School of Salamanca and which characterized by upholding coherent and rigorous positions in respecting the legal demands of the monetary-bank-deposit contract and by being, in general, very critical and distrustful of the practice of banking activities.

³⁴Martín de Azpilcueta, *Comentario Resolutorio de Cambios*, pp. 74–75; emphasis added.

³⁵Not even in the most brilliant and recent work by Murray N. Rothbard, *Economic Thought before Adam Smith*, vol. 1, to which the present article should be considered a humble addendum.

In opposition to this first group of theorists, a second group of “members” (most of them Jesuits) of the School of Salamanca can be clearly distinguished. This group would be led by Luis de Molina and also included Juan de Lugo and, to a lesser extent, Lessius and Domingo de Soto. These authors followed the leadership of Molina and, as we have already explained, are characterized by the wrong legal foundations which they give to the monetary-bank-deposit contract and by admitting that a fractional reserve be permitted, arguing that, more than a deposit, it is a “loan” contract. This is not the place to reproduce all the arguments against the position of Molina respecting the monetary bank deposit contract which merely repeat an error that, very much influenced by the *depositum confessatum*, had been upheld throughout the Middle Ages by the glossators. What we wish to emphasize here is that this second group of authors of the School of Salamanca was much more “comprehensive” with banking activities and even fully justified the practice outside the framework of traditional legal principles. It is not, therefore, inappropriate to consider this second group of authors to form part of an incipient *banking school* within the School of Salamanca who, like their successors of the English and continental banking school several centuries later, not only justified the practice of banking based on a fractional reserve, that is to say, violating basic legal principles, but also believed that this had very beneficial effects on the economy.

Although Luis de Molina’s theoretical arguments on the bank deposit contract are a clear regression and cannot be upheld on the basis of traditional legal principles, it is, however, *curious to draw attention to the fact that this author is the first member of the banking school tradition who was capable of realizing that checks and documents ordering payment at sight of specific amounts charged against the deposits fulfilled exactly the same function as cash*. The assertion that it was the theorists of the English banking school who, in the nineteenth century, first generally argued that the demand deposits of the banks formed an integral part of the monetary supply and thus had the same effect on the economy as banknotes is, therefore, incorrect. More than two centuries earlier, Luis de Molina had already shown this idea clearly in Disp. 409 of his *Tratado sobre los Cambios*. Luis de Molina tells us that “the money is paid to the bankers in two ways: one, in cash, handing over the coins to them; and the other, by trade bills or any other bills which are given to them, by virtue of which he who has to pay a bill becomes a debtor to the bank for the amount which the bill indicates will be paid into the account of he who presents the bill at the bank.”³⁶ Specifically, Luis de Molina

³⁶Luis de Molina, *Tratado sobre los Cambios*, p. 145.

refers to the written documents, which he calls in Latin *chirographis pecuniarum*, which were used as payment in the majority of market transactions. Thus, "although many transactions are made in cash, the majority are made by written documents which evidence either that the bank owes to them or that it agrees to pay, and the money remains in the bank." Molina also says that these cheques are drawn with "sight" or demand value and adds that "these payments are usually called 'sight' because the money must be paid at the moment the bill is presented and read."³⁷

But most importantly Molina expressed, much earlier than Pennington in 1826,³⁸ the essential idea that the total volume of monetary transactions carried out at a market could not be paid with the amount of cash which changed hands there, if it were not for the use of the money which the banks *generate* by their deposit entries and the issue of *checks* against them by the depositors. So that, thanks to the financial activity of the banks, a new amount of money is created from nowhere in the form of deposits, and is used in the transactions. In fact, Molina expressly tells us that "the majority of the transactions are previously carried out (are formalized) by signed documents; *as money is not so abundant as to be able to buy in cash the enormous quantity of merchandise which is taken there to sell, if they must be paid for in cash, nor to be able to carry out so much business.*"³⁹ Lastly, Molina distinguishes very clearly between those operations which imply the grant of credit, as payment of a debt is temporarily postponed, and those which are carried out paying by check or by charging the amount to a bank account, concluding that "it should be observed that it is not considered that credit is bought if the price is charged to the bank account itself, even if at the time cash is not paid; as the banker will pay in cash the debit balance which exists, at least at the end of the market."⁴⁰

Juan de Lugo follows firmly and absolutely the doctrine of Molina and erroneously considers, in the same way as the latter, that the monetary bank deposit is a "loan" which permits that, until the depositors require it, it may be used for the banker's private business.⁴¹

³⁷Ibid., p. 146.

³⁸See James Pennington's memo dated February 13, 1826 "On the Private Banking Establishments of the Metropolis," included as an Appendix in Thomas Tooke, *A Letter to Lord Grenville; On the Effects Ascribed to the Resumption of Cash Payments on the Value of the Currency* (London: John Murray, 1826); Murray N. Rothbard, *Classical Economics: An Austrian Perspective on the History of Economic Thought* (Aldershot, England: Edward Elgar, 1995), vol. 2, pp. 230–33; and F.A. Hayek, "The Dispute Between the Currency School and the Banking School, 1821–1848," Chap. 12 of *The Trend of Economic Thinking: Essays on Political Economists and Economic History*, W.W. Bartley, III, and Stephen Kresge, eds., vol. 3 of *The Collected Works of F.A. Hayek* (London: Routledge, 1991), p. 224.

³⁹Luis de Molina, *Tratado sobre los Cambios*, p. 147

⁴⁰Ibid., p. 149.

⁴¹"De Cambiis," R. P. Joannis de Lugo Hispalensis, Societatis Iesu *Disputationum de Iustitia et Iure, Tomus Secundus, Lugduni 1642*, p. 406, Section 5, No. 60.

Molina and Lugo uphold such a confused position in respect to their legal foundations for the bank deposit contract that they even admit that the contract may simultaneously (!) have a different legal nature, depending on the party under consideration (in other words, it may be a deposit for the depositor and a loan contract for the depositary banker). Apparently, they do not see any incongruence in this position and, as we know, in respect of the bankers' activity, they fix only one limit: that they should act with "prudence" so that, by virtue of the law of large numbers, they always have sufficient liquidity to allow the return of the deposits which are "normally" demanded from them. They do not realize that the criterion of *prudence* which they declare, is not an objective criterion that can guide the bankers' actions. It evidently does not coincide with the capacity to return the deposits held at any given moment and they themselves take great care to emphasize that the bankers are in "mortal sin" when they employ the funds of their depositors in imprudent and speculative activities, *even if they have a favorable result and they are able to return the money to the depositors on time*.⁴² Moreover, the criterion of prudence is not, in itself, sufficient: one can be very prudent and, however, not be very shrewd or even have bad luck in business, so that, when the moment arrives, sufficient liquidity is not available and the deposits cannot be returned.⁴³ What, then, does the criterion of prudence consist of? It is clear that there is no objective reply to this question which could serve as a guide for the bankers' activities. Particularly when the law of large numbers is not applicable to fractional-reserve banking, since the credit expansion which it causes leads to the generation of recurrent cycles of boom and depression which, inevitably, place the bankers in difficult situations. And the fact is that fractional-reserve banking itself, as shown by the Austrian theory of the business cycle, generates liquidity crises and, therefore, the generalized insolvency of the banks. In any case, at the moment of the crisis, it is very possible that the bank cannot pay, in other words that it must suspend payments, and, even if all its creditors are lucky enough to finally collect their money, this will only happen, in the best of circumstances, after a long period of liquidation during which the role of the depositors will change, as they will lose the immediate availability of their money and become *compulsory savers*, who will be obliged to postpone collection of their deposits until the moment when the orderly liquidation of the bank culminates.

⁴²Ibid., pp. 406–7.

⁴³That is to say, in the terminology of Israel M. Kirzner, "Economics and Error," in *Perception, Opportunity and Profit* (Chicago: University of Chicago Press, 1979), pp. 120–36, committing sheer or pure entrepreneurial error (which cannot be insured by the law of large numbers) which causes serious entrepreneurial losses, regardless of the degree of prudence with which one has acted.

The above considerations are those which, without any doubt, lead Tomás de Mercado to indicate that the principles of prudence declared by Molina and Juan de Lugo constitute an objective which, in practice, no banker can meet. It seems as if Tomás de Mercado was aware that such principles did not serve as a practical guide in order to guarantee the solvency of the banks. And, if such principles are not efficient in permanently attaining the objective of solvency and liquidity, it is evident that a fractional-reserve banking system will not be able to meet its commitments under all conceivable circumstances.

The Contemporary Positions of the Jesuits Bernard W. Dempsey and Francisco Belda

In the present century, two Jesuit economists have again studied the doctrine of the scholastics concerning banking, one from the standpoint of the banking school and the other from the position of the currency school. The former is the Spanish Jesuit Francisco Belda, author of an interesting work entitled “*Ética de la creación de créditos según la doctrina de Molina, Lesio y Lugo*” (“Ethics of the creation of credits according to the doctrine of Molina, Lessius and Lugo”).⁴⁴ In fact, for Father Belda it is evident that “from Molina’s description, it may be deduced that, in the case of the bankers, there is authentic credit creation. Thanks to the intervention of the banks, a new purchasing power has been created, which did not exist previously. The same money is used simultaneously twice; the bank uses it for its business and so does the depositor. The overall result is that the payment means that circulation is several times greater than the real amount of money in cash which was originated by them and the bank benefits from all these transactions.” Moreover, Belda considers that, for Molina, “it is licit to do business with the clients’ deposits, provided this is done prudently, not risking being unable to meet one’s obligations on a timely basis.”⁴⁵

With regard to Juan de Lugo, Belda indicates that he gives “a meticulous description of the bankers’ practices. Here, there is explicit approval of credit creation, although not with the formal appearance of *created* credit. The banks do business with the deposits of their clients who, in turn, are not denied the use of their own money. There is an expansion of the payment means, produced by the banks through credits, the discount of trade bills and other economic activities carried on with the money of others. The final result is an increase in the purchasing

⁴⁴Published by the journal *Pensamiento: revista trimestral de investigación e información filosófica publicada por las Facultades de Filosofía de las Compañías de Jesús en España* 73, vol. 19 (January–March 1963): 64–89.

⁴⁵*Ibid.*, pp. 63 and 69.

power in the market very much greater than the amount represented by the cash deposits from which it originates.”⁴⁶

It is evident that Belda is correct in indicating how the doctrines of Molina and Lugo are, from among those of the scholastics, the most favorable towards the banking business. However, we are obliged to criticize Father Belda for not even mentioning the positions of the other members of the School of Salamanca, specifically those of Tomás de Mercado and, above all, of Martín de Azpilcueta and Saravia de la Calle, which are much more rigorous and critical when analyzing the banking institution. Moreover, Belda’s analysis of the contributions of Molina and Lugo is based on a Keynesian conception of the economy, which not only ignores all the negative effects which credit expansion provokes in the structure of production, but also considers it to be highly beneficial to the extent that it increases the “effective demand” and national income. Belda’s analysis is, therefore, a study of the contributions of the members of the School of Salamanca from the point of view of the Keynesian and banking schools and is extremely confused regarding the legal justification of the institution of the monetary bank deposit, tending, therefore, to consider fractional-reserve banking to be legitimate.

There is, however, an economic treatise by another notable Jesuit, Father Bernard W. Dempsey, entitled *Interest and Usury*,⁴⁷ in which he analyzes the position of the members of the School of Salamanca on the banking business employing a profound knowledge of monetary and capital theory, very much superior to that of Father Belda.⁴⁸

Curiously enough, Dempsey develops his thesis, not by analyzing the positions of the theorists of the School of Salamanca who are most unfavorable to banking activity (Saravia de la Calle, Martín de Azpilcueta Navarro and Tomás de Mercado), but by concentrating on the works of the representatives who are closest to the banking school, Luis de Molina, Juan de Lugo and Lessius, making an interpretative

⁴⁶Ibid., p. 87. The reference to Juan de Lugo corresponds to *Disputationum de Iustitia et Iure, Tomus Secundus* pp. 60–62.

⁴⁷Bernard W. Dempsey, *Interest and Usury*, published with an introduction by Joseph A. Schumpeter (Washington, D.C.: American Council of Public Affairs, 1943). Attention should be drawn to the fact that Father Belda’s article arose as a criticism, from the Keynesian point of view, of the theses upheld by Dempsey in this book. I would like to thank Professor James Sadowsky of Fordham University, who provided me with a copy of Dempsey’s book, which was not available in Spain.

⁴⁸Father Dempsey’s broad theoretical knowledge and familiarity with the economic doctrines of Ludwig von Mises, Friedrich A. Hayek, Wicksell, Keynes and others is very much emphasized in Schumpeter’s “Introduction” to his book. Moreover, Schumpeter quotes and praises Dempsey in his *History of Economic Analysis* (New York: Oxford University Press, 1980), pp. 95–96 and 104.

study of the works of these authors which leads him to conclude that *from the point of view of their own doctrines, banking activity based on fractional reserves would not be legitimate*. Dempsey's conclusion is based on the application of the traditional principles on usury defended by these authors to the banking institution and the economic effects thereof which, although they were completely unknown at the time the School of Salamanca was writing, had, however, already been theoretically revealed by Mises and Hayek when Dempsey wrote his book. In fact, although Molina and Lugo's more favorable treatment of banking must be acknowledged, Dempsey expressly indicates that the loans which are created from nowhere by the banks, as a result of practicing their activity with a fractional reserve, means the generation of purchasing power which does not require any prior saving or sacrifice and gives rise to important damage to a large number of third parties, who see how the purchasing power of their monetary units decreases as a consequence of the inflationary credit expansion of the banks.⁴⁹ According to Dempsey, this creation of purchasing power from nowhere, which does not imply any prior loss of other people's purchasing power, is contrary to essential legal principles, as constructed by Molina and Lugo themselves and, in this respect, should be condemned. Specifically, Dempsey affirms that "we may conclude from this that a Scholastic of the seventeenth century viewing the modern monetary problems would readily favor a 100-percent-reserve plan, or a time limit on the validity of money. A fixed money supply, or a supply altered only in accord with objective and calculated criteria, is a necessary condition to a meaningful just price of money."⁵⁰

⁴⁹"The credit expansion results in the depreciation of whatever circulating medium the bank deals in. Prices rise; the asset appreciates. *The bank absolves its debt by paying out on the deposit a currency of lesser value . . .* No single person perhaps would be convinced by a Scholastic author of the sin of usury. But the process has operated usuriously; again we meet systematic or institutional usury . . . The modern situation to which theorists have applied the concepts of divergence of natural and money interest, divergences of saving an investment, divergences of income disposition from tenable patterns by involuntary displacements, all these have a sufficient common ground with late medieval analysis to warrant the expression 'institutional usury' for the movement heretofore described in the above expressions." Father Bernard W. Dempsey, *Interest and Usury*, pp. 225 and 227-28.

⁵⁰*Ibid.*, p. 210. Incidentally, Father Dempsey points out that the theory of time preference may even date from Saint Thomas Aquinas, as it was expressly stated by one of the latter's most brilliant pupils, Giles Lessines, for whom "future goods are not valued so highly as the same goods available at an immediate moment of time, nor do they allow their owners to achieve the same utility. For this reason, it must be considered that they have a more reduced value in accordance with justice." See p. 426 of *Opusculum LXVI, De usuris in communi et de usurarum contractibus*, written by Aegidius Lessines in 1285 (quoted by Bernard Dempsey in Note 31 of p. 214). Dempsey's discovery of Lessines exposition of time preference was not included in Murray N. Rothbard's *Economic Thought before Adam Smith*, in which Rothbard considers San Bernardino of Siena and Conrad Summenhart to have been in 1431 and 1499 the first expositors of time preference theory (pp. 85 and 92).

Dempsey states that the credit expansion generated by the banking industry tends to depreciate the purchasing power of money, so that the banks tend to return the monetary deposits claimed from them in monetary units with an increasingly reduced purchasing power.⁵¹ He therefore correctly concludes that, if the members of the School of Salamanca had had a detailed theoretical knowledge of the functioning and implications of the economic process to which fractional-reserve banking gives rise, it would have been described as a perverse, vast and illegitimate process of *institutional usury*, even by Molina, Lessius, and Lugo themselves.

⁵¹This is the same argument given by the great libertarian Jesuit Juan de Mariana in his book *De monetae mutatione (On the Alteration of Money)* published in 1609. Mariana condemns as robbery any government debasement of coins whereas Dempsey follows the same reasoning in relation to the even more disturbing credit inflation created by banks. On Juan de Mariana see the most brilliant analysis of Murray N. Rothbard, *Economic Thought before Adam Smith*, p. 119.

In Defense of Fiduciary Media—or, We are *Not* Devo(lutionists), We are Misesians!¹

George Selgin and Lawrence H. White

The Murray Rothbard both of us knew was committed to a frank and vigorous contest of ideas. He understood that an expression of disagreement was not an expression of disrespect—quite the contrary. Here we wish to honor Rothbard’s memory by addressing a set of issues surrounding fractional-reserve banking, issues on which we disagree with some of Rothbard’s conclusions despite beginning (we believe) from many of the same premises. Our main concern is to defend the freedoms to issue and use fiduciary media of exchange. The vehicle for our defense is a response to criticisms of our views by Hans-Herman Hoppe in his article “How is Fiat Money Possible?—or, The Devolution of Money and Credit” (1994). Subsequent to Hoppe’s article, Jesús Huerta de Soto (1995) and Jörg Guido Hülsmann (1996) have also offered criticisms of our position in lengthy articles in this journal. We address at a few points in the text below what we take to be de Soto’s main arguments. Hülsmann’s article has appeared too recently for us to address it directly here, but its arguments closely parallel Hoppe’s. In particular, Hülsmann, like Hoppe, fails to appreciate Mises’s (fairly standard) explanation of why fractional-reserve banking is feasible. We therefore believe that our rebuttal to Hoppe serves to rebut Hülsmann’s main arguments as well.

The Origins of Fiat Money

It should be understood at the outset that fiduciary media, i.e., demandable bank claims that are not 100 percent backed by bank

*George Selgin and Lawrence H. White are associate professors of economics at the University of Georgia.

¹With apologies to Devo, the ‘80s rock band who used the slogan “Are We Not Men? We Are Devo!”

The Review of Austrian Economics Vol. 9, No. 2 (1996): 83–107
ISSN 0889-3047

reserves of basic money, are *not* a type of fiat money. We do not intend to defend fiat money here, and have not defended it in our previous writings. Professor Hoppe unfortunately suggests otherwise. In the course of arguing that “no fiat money can ever arise ‘innocently,’” i.e., purely from free-market forces rather than from government intervention, Hoppe (1994, p. 49) criticizes at length what he calls “various prominent counterarguments.” He names us as authors of one supposed counterargument, as though we had argued for the possibility of an innocent fiat money. In fact we have explicitly argued the opposite. In discussing the institutional evolution of free markets in money and banking we concluded (Selgin and White 1987, pp. 453–4) that “commodity-based money would persist in the absence of intervention, for the reason that the supreme salability of the particular money good is self-reinforcing,” and that there is thus “no basis for the spontaneous emergence of a multi-commodity standard or of any pure fiat standard.”²

How then do we think fiat money came to be? Our writings on the question have been plain enough. White (1989, pp. 59–61) has answered that “government has suppressed commodity money in favor of fiat money” and added: “I do not know of a single historical case of fiat money supplanting commodity money through competition rather than compulsion. . . . Historically, the introduction of fiat money . . . has come about by the permanent suspension of redeemability of the central bank’s liabilities, enriching only the government.” Selgin (1994c, p. 811) has addressed the question at length,³ affirming the conclusion reached by Mises (and by Rothbard) that “States have never established fiat monies through ‘social compacts,’ . . . but rather have had to create them at first by taking convertible commodity-based monies that were already in circulation and ‘depriving them of their essential characteristic of permanent convertibility.’” (The first internal quotation is from Rothbard, the second from Mises.)

The factual origins of fiat money are thus not, in our view, to be found in the free market. But is fiat money nevertheless a desirable innovation? We have not said so, and we do not think so. We regard the dismantling of commodity standards by governments as a great tragedy, something accomplished by highly objectionable means and having economically destructive consequences. The central banks’ devaluation and finally repudiation of their contractual obligations to redeem their notes and deposits in gold involved massive confiscations of private wealth, and paved the way for ruinous episodes of inflation and depression the likes of which would not have been experienced under an unmanaged commodity standard.

²This essay is reprinted in White (1989), a book Hoppe cites.

³Selgin’s paper was available at the time Hoppe wrote, having been presented at a Mises Institute conference which he attended in 1992.

“Fiat” Redefined by Fiat

Hoppe’s grouping us with defenders of fiat money is therefore puzzling, especially given that he recognizes (pp. 69–70) that our monetary ideal “is a universal commodity money such as an international gold standard.” So how are we supposed to favor both fiat and commodity money? The answer lies in a verbal sleight-of-hand. Although beginning his article with what seems to be the conventional definition of fiat money (“a medium of exchange which is neither a commercial commodity, a consumer, or a producer good, nor title to any such commodity; i.e., irredeemable paper money”), Hoppe tacitly redefines the category of fiat money to include banknotes and deposits that are redeemable-on-demand claims to commodity money, so long as they are not backed 100 percent by reserves of commodity money.⁴ It is true that we have offered both ethical and economic arguments in defense of the contractual practice of fractional-reserve banking.

Any author is free to redefine terms as he pleases, but it is misleading for him to depart from an established usage without announcing plainly that he is doing so. Hoppe’s expanded usage of “fiat money” is unorthodox, to say the least, even from an Austrian point of view. Mises (1966, p. 429, emphasis added), for one, defined fiat money as “money consisting of mere tokens which can neither be employed for any industrial purpose *nor convey a claim against anybody*.” He carefully distinguished the category of base money or “money in the narrower sense,” which includes gold coins (in a gold standard regime) and true fiat currency (in a fiat money regime), from the category of “money substitutes,” which includes fractionally-backed checking deposits and banknotes (which of course *do* convey a claim against banks). Finally, Mises (1966, p. 433; 1980, appendix B) referred to that portion of redeemable money substitutes backed by assets other than base money as “fiduciary media,” not as any kind of fiat money. Rothbard (1970a, p. 703) follows Mises’s terminology in every particular. According to the Misesian terminology, then, a fractionally-backed banknote that is *de facto* redeemable, and is recognized by the public to be redeemable, is not an example of fiat money. Contrary to Hoppe’s (pp. 49, 73) innovative phraseology, it is neither a “fractional” fiat money nor a “partial” fiat money.⁵ It is instead a fractionally or partially fiduciary medium.

⁴Perhaps his view is that, even when in practice a fractional-reserve bank for years fulfills every redemption request that actually comes to it, nonetheless its notes should *really* be considered *irredeemable* because the bank *would* default if all its notes and demand deposits were presented for redemption simultaneously. And for the same reason Hoppe may view the title conveyed by a banknote’s contractual pledge that the bank “will pay to the bearer on demand” as not *genuinely* a title at all.

⁵Redeemable bank liabilities are not fiat money even if the (fractional) bank reserves

Labels aside, Hoppe's lumping together of fiduciary media with fiat money is substantively misleading, because it blurs important theoretical differences between the two. The determinants of the quantity of fiduciary media in a fractional-reserve banking system are quite distinct from the determinants of the quantity of fiat money. Economic factors strictly limit the quantity of fiduciary media a banking system can issue, given its reserves of base money. The quantity of fiat money, by contrast, is not subject to any natural economic limit.⁶ We have argued (Selgin and White 1994, pp. 1734–5) that a natural limit to the quantity of fiat-type (i.e., irredeemable, non-commodity) money would be lacking even if such money were issued by competing firms. Thus Hayek's (1978) proposal for private fiat-type money unfortunately fails to secure the quantity and value of money. A "free banking" regime with competing issuers of redeemable notes and deposits is quite distinct from a Hayekian regime of "competing fiat monies."

Normative and Positive Questions

Given the difference between fiduciary media and fiat money, as those terms are used by Mises and Rothbard, the questions arise whether it is ethically or economically defensible to allow *fiduciary media* to be issued. We side with Mises, and part company from Rothbard and Hoppe, by acknowledging the legitimacy and practical advantages of fiduciary media and fractional-reserve banking. We base the legitimacy argument on Rothbardian normative analysis, and the practical-advantages argument on Misesian economic analysis.

Rebutting the Charge of Fraud

Rothbard (1962, 1983b, 1990, 1995) long argued that fractional-reserve banking is inherently fraudulent, and Hoppe follows Rothbard down this unfortunate blind alley. We find the inherent-fraud position impossible to reconcile with Rothbard's (1983a, pp. 133–48) own title-transfer theory of contract, which we accept, and which Rothbard otherwise uses to defend the freedom of mutually consenting individuals to engage in capitalist acts with their (justly owned) property. Rothbard (1983a, p. 142) defines fraud as "failure to fulfill a voluntarily-agreed

themselves consist of fiat money. In Misesian terms, a bank-issued medium of exchange is a "money substitute," i.e., a substitute for money proper (either for fiat or for commodity money).

⁶To be precise, we mean the quantity measured in units of account, holding the definition of the unit of account constant.

upon transfer of property.”⁷ Fractional-reserve banking arrangements cannot then be *inherently* or *inescapably* fraudulent. Whether a particular bank is committing a fraud by holding fractional reserves must depend on the terms of the title-transfer agreements between the bank and its customers.

Rothbard (1983a, p. 142) in *The Ethics of Liberty* gives two examples of fraud, both involving blatant misrepresentations (in one, “A sells B a package which A says contains a radio, and it contains only a pile of scrap metal”). He concludes that “if the entity is not as the seller describes, then fraud and hence implicit theft has taken place.” The consistent application of this view to banking would find that it is fraudulent for a bank to hold fractional reserves if and only if the bank misrepresents itself as holding 100 percent reserves, or if the contract expressly calls for the holding of 100 percent reserves.⁸ If a bank does not represent or expressly oblige itself to hold 100 percent reserves, then fractional reserves do not violate the contractual agreement between the bank and its customer (White 1989, pp. 156–57). (Failure in practice to satisfy a redemption request that the bank is contractually obligated to satisfy *does* of course constitute a breach of contract.) Outlawing voluntary contractual arrangements that permit fractional reserve-holding is thus an intervention into the market, a restriction on the freedom of contract which is an essential aspect of private property rights.

Hoppe declares our defense of the freedom to make fractional-reserve-compatible contracts to be “silly” because, he asserts, “few if any” depositors have ever realized that some of their deposits are being loaned out, even though (as he acknowledges) the payment of interest on deposits would otherwise be impossible. We doubt that most depositors are as naive as Hoppe believes. As Rothbard (1990, p. 47) has correctly observed, “It is well-known that banks have rarely stayed on a ‘100 percent’ basis very long.” We thus find it hard to believe that most people who patronize fractional-reserve banks do so under the delusion that 100 percent of the money they deposit remains in the

⁷A more standard definition of fraud confines it to a *willful* or *deliberate* deception for purposes of gain. Thus an unintended failure to meet the terms of an agreed transfer due to unexpected circumstances beyond the party’s control, would constitute a breach of contract, but not a fraud. Nothing herein turns on this distinction, though.

⁸Whether it is fraudulent to hold fractional reserves against a bank liability does not depend *per se* on whether it is a demand or time liability, but only on whether the bank has misrepresented itself as holding 100 percent reserves. The demandability of a particular claim issued by a bank, i.e., the holder’s contractual option to redeem it at any time, is not *per se* a representation that the bank is holding 100 percent reserves against the total of its demandable claims. Rothbard (1990, pp. 49–50) argues otherwise, based on the view that a bank’s demand deposits and notes are necessarily “warehouse receipts” and not debts. We do not see why bank and customer cannot contractually agree to make them debts and not warehouse receipts, and we believe that historically they have so agreed.

bank's vault until the moment they ask for it back. (We return to this issue below.)

But whether the informed would-be customers of fractional-reserve banks be a majority or a minority, their freedom of contract is at stake. If *any* person knowingly prefers to put money into an (interest-bearing) fractional-reserve account, rather than into a (storage-fee-charging) 100 percent reserve account, then a blanket prohibition on fractional-reserve banking by force of law is a binding legal restriction on freedom of contract in the market for banking services.

Walter Block (1988, pp. 28–30), though he (following Rothbard) judges fractional-reserve banking “as presently constituted” to be “a fraud and a sham,” acknowledges that fractional-reserve banking *could be* non-deceptive and voluntary. To make it so, Block argues, the bank needs to affix an adequate disclaimer to banknotes and deposit contracts regarding the bank's fractional-reserve-holding and redemption policies. Hoppe (1994, p. 71), citing Block, similarly allows that fractional-reserve practices would be non-fraudulent if the bank explicitly informed depositors that it reserved the right to “suspend or defer redemption” at any time.

If the proponents of the “fraud” objection to fractional-reserve banking thus concede that the objection vanishes when banks apply the equivalent of a “warning sticker,” then they concede that fractional-reserve banking is not inherently fraudulent. Fraud occurs only if a bank's customers are misled about its practices. The remaining normative debate boils down to the question of whether a warning sticker really is needed to avoid misleading customers (which in our view depends on whether the reasonable default assumption, absent a sticker, is really that 100 percent reserves *are* being held), and, if so, to the question of how explicit the sticker must be. There is also the positive question of whether fractional-reserve banknotes and deposits really could circulate among an informed public.

Our view is that a mandatory “warning sticker” is certainly less objectionable than an outright ban on fractional-reserve banking, and would not impede the practice of fractional-reserve banking, but that it is not really needed to avoid misrepresentation, because a “deposit” is *not* commonly understood to be a 100-percent-reserve bailment unless otherwise specified. As Rothbard (1970b, p. 34) once described the libertarian approach to preventing product adulteration, “if a man simply sells what he calls ‘bread,’ it must meet the *common definition* of bread held by consumers, and not some arbitrary specification. However, if he *specifies* the composition on the loaf [Rothbard does not suggest that this should be mandatory], he is liable for prosecution if he is lying.” We maintain that the common definition or default meaning of a “bank deposit” is, as

courts have recognized (Rothbard 1983b, pp. 93–94), that of a debt claim against the bank and not of a warehouse receipt.

Block and Hoppe propose slightly different warnings as adequate to avoid fraud. It is not clear whether they are merely offering examples, or instead believe these to be the only sorts of adequate warnings. Block's warning would detail the bank's reserve ratio and its policy for meeting redemptions when they exceed its reserves (e.g., first-come first-served). His example seems to assume that the bank would hold a *fixed* reserve ratio (because it specifies the precise ratio on its notes). The bank and its customers might well both prefer, however, to allow the bank discretion to vary the ratio as prudence dictates. Under varying conditions, a varying ratio is necessary to maintain a constant default risk. Hoppe's warning would inform claim-holders that the bank reserves the right to suspend or defer redemption at any time.⁹ But some banks and their customers might prefer a demandable debt contract that does not give the bank any such right to suspend. What then?

Hoppe likens his warning to the "option clauses" historically placed on banknotes, but it should be noted that such clauses only allowed for the deferral, or temporary suspension, and never for the indefinite suspension of redemption (who, after all, would freely choose to take a *permanently* suspendable note?). The Scottish banks that issued option-clause notes explicitly reserved the right to defer redemption for a specified period, in which case the note would be repaid with a specified (and high) interest bonus.¹⁰ In practice the banks went decades without invoking the option, and the clause-laden notes circulated easily at par, because the banks were not expected to invoke the option. Hoppe's prediction that option-clause notes "would be uniquely *unsuited* to serve as a medium of exchange" is false, to judge by the Scottish evidence.

Equally without historical support is Block's (1988, pp. 30–31) suggestion that, because the holder of a note issued by a bank with a 20

⁹Hoppe would also have the bank inform its borrowers that their loans can be recalled at any time. On this odd suggestion see footnote 13 below.

¹⁰Checkland (1975, p. 67) provides a specimen of an optional note issued by the Royal Bank of Scotland. The face of the note reads, in fairly large print (occupying practically the entire face), "The Royal Bank of Scotland . . . is hereby obliged to pay to [name] Or the Bearer, One Pound Sterling on demand Or, in the Option of the Directors, One pound Six pence Sterling at the End of Six Months after the day of the demand & for ascertaining the demand & Option of the Directors, the Accomptant & One of the Tellers of the Bank are hereby ordered to Mark & Sign this Note on the back of the same." The Bank of Scotland, also known as "the Old Bank," introduced the option clause in 1730. Checkland (1975, p. 68) comments that "The adoption of the clause does not seem to have impaired the Old Bank's note issue." The public presumably realized that the bank would try to avoid having to invoke the option to defer redemption, both for reputational reasons and because the bank would then, under the terms of the clause, have to pay interest on its notes. The bank did not in fact invoke the option until 1762. Option clauses were outlawed in 1765.

percent reserve has only a 20 percent chance of redeeming it in the event of a bank run, a note issued by a bank known to hold fractional reserves is indistinguishable from a lottery ticket, and would be valued below par if the public were to “fully digest” the implications of its issuer’s fractional reserves. It is true that a particular bank’s notes would be valued below par if market participants worried that they might not be able to redeem the notes ahead of an imminent run on that bank. But such notes, on which default was considered a non-negligible risk, would not continue circulating, even at a discount. They would immediately be presented for redemption, and thus removed from circulation. The surviving brands of notes would be only those for which all redemption demands made in practice were expected to be met (see Mises 1966, p. 445). Fractional-reserve notes issued by respected banks—and such banks were not historically rare—were in fact able to circulate widely *at face value* because other banks and the public rightly recognized that the practical likelihood of experiencing any difficulty in redeeming the notes was negligibly small.

The notion that a fractionally-backed banknote is akin to a lottery ticket seems to rest on a failure to appreciate the simple fact that fractional-reserve banking *is feasible*, that is, that a fractional-reserve bank can in practice continually fulfill its contractual obligation to redeem on demand. A fractionally-backed claim to basic money, a banknote or checking deposit balance, can itself serve as a medium of exchange. Because it is thus useful even without being redeemed for basic money, there is no reason to expect all the claims issued by a bank (unlike claims to bread, or winning claims against a lottery) to be redeemed in a given period. As Mises (1980, pp. 299–300) put it, a banker “is therefore in a position to undertake greater obligations than he would ever be able to fulfill; it is enough if he takes sufficient precaution to ensure his ability to satisfy promptly that proportion of claims that is actually enforced against him.”

A demand deposit is the limiting case of a short-term deposit. Hoppe’s view that it is infeasible for a bank to hold a fractional reserve against its demand liabilities would seem to imply that it is generally infeasible for a bank to borrow short and lend long, or to practice anything less than perfect maturity matching of liabilities with assets. Rothbard (1983, p. 99) argues explicitly that any bank that practices maturity-mismatching, i.e., has time deposits coming due before loan pay-offs arrive, is violating “a crucial rule of sound financial management.” The practice is feasible (does not inevitably doom a bank), however, if the bank can count on rolling over or replacing at least some of its time deposits as they come due. Maturity-mismatching clearly does involve risks: not only liquidity risk, but also interest-rate risk. But surely the

rules of sound financial management do not say that risk should never, ever be taken. Rather, they call for risk to be balanced against the return for risk-taking. A risk can be worth taking if the risk is small enough relative to the reward for taking it. When long-term market interest rates are higher than short-term rates, banks do earn a reward for assuming the risks involved in intermediating short-term deposits (including demand deposits) into longer-term loans. The view that fractional-reserve banking and maturity mismatching in general are “inherently unsound” practices seems to suggest that no bank should ever knowingly engage in any risk-return tradeoff with regard to the maturity structure of its balance sheet.

Jesús Huerta de Soto (1995, p. 30) rejects “the trite argument that the ‘law of large numbers’ allows the banks to act safely with a fractional reserve,” on the grounds that “the degree of probability of an untypical withdrawal of deposits is not, in view of its own nature, an insurable risk.” It is true that the atypical withdrawals known as bank runs are not random events. But it does not follow that a bank cannot survive with fractional reserves, because solvent banks are not inherently run-prone. Even in countries (e.g., Scotland, Sweden, Canada) where the legal system vigorously enforced the banks’ contractual obligation to pay on demand (and even where legislatures outlawed the contractual escape hatch from runs provided by an option clause), well-known banks with fractional reserves did not experience runs and continually met all their redemption demands for decades (Dowd 1992; Selgin 1994a).

If runs *were* a problem even with solvent banks—that is, if depositors ran simply out of fear that others would run, thereby forcing any less-than-perfectly-liquid bank to default—an option clause would be an available contractual remedy.¹¹ An option clause in note and demand deposit contracts gives the bank the option to suspend payments in the event of a run, for a period long enough to allow the bank to liquidate its non-reserve assets in orderly fashion. To make the clause acceptable to customers, judging by the historical example of the Scottish optional notes, the bank would have to specify the period of suspension (or at least its maximum length), and obligate itself to make a compensatory interest payment (in addition to returning the note’s face value in base money) at the end of any suspension period. This

¹¹It is in this connection, and not in connection with the “fraud” issue, contrary to Hoppe’s account of our argument (1994, p. 71), that we consider the option clause important. But we can see that from Hoppe’s viewpoint the clause also eliminates the charge of fraud, since the bank is no longer promising *unconditionally* to redeem its claims on demand, and therefore the total of its *unconditionally* demandable claims no longer exceeds its reserves.

payment would not only compensate the customer for the inconvenience and delay, but would also give the bank a visible incentive not to invoke the option except when necessary (in technical jargon, it would make the contract “incentive-compatible”; it avoids a potential moral hazard problem by penalizing a bank that skimps on reserves and thereby runs too great a risk of suspension). Historically, as discussed in the text, some banks did write such option-clause contracts, where their legislatures did not forbid them to do so.

But how do we know that not everyone who accepted a fractional-reserve note at face value was in the dark about its fractional backing? At the very least we know that competing *banks* participated in clearing arrangements in which they agreed to accept one another’s notes at par. Certainly the bankers were not in the dark. They did not expect—or find—defaults at the clearinghouse to be more than extremely rare.

Third-Party Effects

Apart from the fraud and feasibility questions, Hoppe (pp. 70–71) offers another (“and more decisive”) set of reasons why fractional-reserve banking contracts should be banned: they have spill-over effects on others. His argument bears quoting:

Whenever a bank loans its “excess” reserves to a borrower, such a bilateral contract affects the property of third parties in a threefold way. First, by thereby increasing the money supply, the purchasing power of all other money owners is reduced; second, all depositors are harmed because the likelihood of their successfully recovering their own possessions is lowered; and third, all other borrowers—borrowers of commodity credit—are harmed because the injection of fiduciary credit impairs the safety of the entire credit structure and increases the risk of a business failure for every investor of commodity credit.

Let us consider these three third-party effects in turn.

(1) The first effect, the reduction in the purchasing power of money, provides no justification for legally barring the bank’s action. To think that it does is to commit the elementary mistake of confusing spill-overs from others’ actions to the *value* of C’s property, which are an inescapable free-market phenomenon and not a violation of C’s property rights, with *physical invasions* of C’s property, which are of course inconsistent with the protection of C’s property rights.¹² It should be obvious that if A and

¹²Economists conventionally distinguish a “pecuniary externality,” an effect on someone’s wealth transmitted via the price system, from a “technological externality,” a physical or otherwise direct interference with someone’s consumption or production.

B are to be barred from any transaction that merely affects the *market value* of C's possessions, without any physical aggression or threat against C or C's rightful property, then the principles of private property, freedom of contract, and free-market competition are completely obliterated. Is B to be barred from offering to sell compact disc recordings to A, merely because doing so reduces the market value of C's inventory of vinyl records?

To further illustrate the point, consider another non-banking example. Suppose that A, who owns but seldom uses a Florida condominium, contracts with B to time-share B's condominium. A then sells his own condominium, causing the value of neighbor C's condominium to fall. Does this mean that the contract between A and B should not be allowed? Has A robbed C? Not according to the Rothbardian view of property rights. If Rothbard's view of property rights is accepted, Hoppe's first effect is invalid as a ground for thinking that the principle of freedom of contract excludes fractional-reserve contracts.

(2) Hoppe's second supposed effect is that all depositors are "harmed" by the bank lending out any of its reserves, because the likelihood of their successfully redeeming their own deposits is lowered. But if those depositors have freely and knowingly agreed to fractional-reserve contracts, rather than choosing to store their money in a 100-percent-reserve institution, they have agreed to take the risk. Presumably they have agreed in order to get the deposit interest payments (or unpriced bank services) that the revenue from bank lending makes possible, and which competition for depositors compels the bank to provide to its customers. By the principle of demonstrated preference (Rothbard 1957) depositors must be presumed to benefit from the package they have agreed to accept, risk and all.

(3) Finally, Hoppe's claim that "fiduciary credit impairs the safety of the entire credit structure" is difficult to evaluate, because Hoppe does not explain how this effect is supposed to work.¹³ We imagine that Hoppe

The first is an interdependence through the market; the second is an interaction outside the market.

De Soto (1995, p. 33) fails to grasp this distinction when he mischaracterizes the pecuniary externality from fiduciary media as a "tragedy of the commons," a term that properly applies only to a particular sort of technological externality.

¹³In one passage Hoppe (p. 70) remarks that fractional-reserve banks did not "inform that some or all of the credit granted to them had been created out of thin air and was subject to being recalled at any time," and he proposes that a non-fraudulent fractional-reserve bank would have to warn borrowers "that their loans may be instantly recalled." Perhaps Hoppe believes that fractional-reserve banks typically have a secret right to recall their loans at any time, and perhaps this underlies his belief that their loans make the credit structure riskier. But we are baffled as to where he might have gotten such an unfounded idea. Fractional-reserve banks do not have the option to call in loans except where the option is explicitly specified in the loan contract.

has in mind something like the notion Adam Smith (1981, p. 321) expressed by saying that “The commerce and industry of the country . . . though they may be somewhat augmented [because less of the country’s capital stock is being tied up in gold and silver], cannot be altogether so secure, when they are thus, as it were, suspended upon the Daedalian wings of [bank-issued] paper money, as when they travel about upon the solid ground of gold and silver.” If so, we grant the point that a risk to a bank and its customers is involved in the bank’s funding loans by issuing banknotes and demand deposits, rather than relying entirely on time deposits. There may even be spill-over effects upon the risks faced by third parties. Nonetheless we side with Smith in thinking that the risks are small in comparison with the benefits. Benefits accrue to bank depositors and note-holders, who receive interest and services paid for by the extra bank revenue generated from lending out a portion of its liabilities. Benefits accrue to bank borrowers who enjoy a more ample supply of intermediated credit, and to everyone who works with the economy’s consequently larger stock of capital equipment. And benefits must accrue to bank shareholders, who could choose to have the bank not issue demand liabilities if they found the risks not worth bearing.

We consider below the resource cost savings and “inherent instability” of a fractional-reserve system. With both factors considered, a higher standard of living is made possible by allowing those members of the public who so prefer to substitute fiduciary media for the holding of gold and silver coin (White 1992, pp. 520–21). As Mises (1980, p. 359) put it: “Fiduciary media tap a lucrative source of revenue for their issuer; they enrich both the person that issues them and the community that employs them.”

The entire credit structure can be made radically unsafe by central banking and other government intervention (Selgin 1989; Salsman 1990), but the effects of those measures should not be charged to fractional-reserve banking as such. As we discuss in more detail below, an unhampered fractional-reserve banking system is not inherently unstable or prone to cyclical over-expansion.

When a loan is callable the call provision is thus no secret to the borrower. Historically, call loans have been a very small share of all bank loans.

We also reject the notion, expressed in the passage quoted above, that *competitive* banks issuing redeemable liabilities can create credit “out of thin air.” By the nature of the balance sheet, all bank loans must be funded by liabilities or equity. Neither source of funds can be conjured out of thin air. No one is forced to hold a competitive bank’s redeemable liabilities or to buy its shares; anyone can hold claims on other banks instead, or on no bank. A competitive bank must therefore *expend real resources to attract a clientele* by the provision of interest and services. The notion that a bank can extend credit costlessly or gratuitously is valid only with respect to the inframarginal credits of a monopoly bank, or to an issuer of a forced tender; it does not apply to a bank in a competitive system (see Mises 1980, pp. 346–7).

The Popularity of Fractional-Reserve Banking

Let us return to the question of how large or small is the pool of voluntary fractional-reserve depositors. The group whose freedom of contract we are concerned with here is not a small eccentric bunch, but is the great mass of people who have demonstrated that they do prefer banks that operate on fractional reserves. To quote Rothbard (1990, p. 47) again, with emphasis added, “It is *well-known* that banks have rarely stayed on a ‘100 percent’ basis very long.”¹⁴ Yet depositors continue to patronize these banks, demonstrating their preference for them.

There are several reasons why fractional-reserve practices are and have been well-known.

First, as Hoppe (p. 70) acknowledges, from the fact that banks pay interest on demand deposits “it *should* have been clear that the bank *must* loan out deposits.”¹⁵ A bank that offers interest on its demand deposits, and does not charge warehousing fees, gives its depositors clear notice that some fraction of their funds will be put to work and not warehoused.

Second, if the vast majority of people thought that their banks held 100 percent reserves, bank runs would have occurred only when there was a suspicion that the banker was about to abscond with the reserves.¹⁶ The history of banking before deposit insurance indicates that when bank runs have occurred, this has typically been for other reasons (Gorton 1988). Depositors’ behavior has generally been consistent with their realizing all along that their banks held fractional reserves, and that they would pay them out on a first-come first-served basis. Generally depositors remained confident that the reserves were sufficient to meet all actual demands for cash. But occasionally, and more frequently in

¹⁴Likewise de Soto (1995, p. 31), who regards the 100 percent reserve custodial deposit as a form consecrated by the Roman Law tradition, and who would (it seems) deny transactors the freedom to make alternative (non-traditional) demand deposit contracts, does at least recognize that modern banks have been “open” about holding fractional reserves.

¹⁵Given his recognition that competitive fractional-reserve banks pass loan revenues on to depositors in the form of interest on demand deposits, we are baffled as to how Hoppe (p. 66) can—in the immediately preceding sentence, no less—claim that fractional-reserve banking “leads to a unilateral income redistribution in the bank’s favor.”

¹⁶It is true that a bank that mixes a time deposit business with its (100 percent reserve) demand-liability business might become insolvent, and might therefore be runnable even without any absconding. But depositors who really want 100 percent-reserve bailment contracts receive no apparent advantage from such a mixture, and they should learn over time to avoid riskier mixed institutions in favor of pure warehouse banks. If such depositors were common the market would enforce the “strict functional separation of loan and deposit banking” that Hoppe (p. 74) wishes to see. With such a separation, the mere fact that a bank offers loans is a clear tip-off that it is not a 100-percent-reserve institution.

some systems than in others, they lost their confidence, and staged runs. Runs were typically triggered by reasonable doubts about a bank's solvency. Heavier government intervention was a background condition explaining why some countries (like the United States) but not others (like Canada) had chronically weak or insolvency-prone banks (Selgin 1994a).

Early in the history of banking there may have been a case of a run being triggered by depositors' sudden realization that their bank held only fractional reserves.¹⁷ But if such a realization had been the typical cause of runs in the nineteenth and twentieth centuries, it would be difficult to explain why runs usually affected only one particular bank or an associated set of banks, and not every single fractional-reserve bank simultaneously. Running depositors who successfully withdrew their money often transferred it to other fractional-reserve banks, thought to be safer, rather than hoarding cash as they would have done if they feared fractional-reserve banks generally (Kaufman 1994). It would be farfetched to account for such behavior by insisting that the depositors had run because they had learned to their horror that their own banks had been holding fractional reserves, but were so naive as to put their money into another set of banks without suspecting them of similar practices.

Third, banks and banking legislation were widely debated in the popular press during the nineteenth century. All discussions we are aware of took it as common knowledge that banks operate on fractional reserves. It would be impossible to think that banks were holding 100 percent reserves after reading in the newspaper about such measures as, for example, the New York State Safety Fund (a deposit insurance scheme), or the so-called "free banking" acts that compelled state-chartered banks to hold specific sorts of interest-bearing assets as collateral against banknote liabilities.¹⁸

¹⁷Some writings suggest that this occurred with the Bank of Amsterdam (Hildreth 1968, p. 12, is a bit vague). But the details behind this story, as presented in Van Dillen (1934), are rather more complicated. First of all, the Bank of Amsterdam was not expressly forbidden to make loans until 1802, and, although it kept close to 100 percent reserves throughout much of its existence, there were long periods (e.g., 1723–1761) when its reserves fell substantially below its deposit balances, the difference consisting of loans made to the East India Company and to the Amsterdam Treasury. The decline in the Bank's reputation in the mid-1780s appeared to reflect not a sudden realization that it held less than 100 percent reserves, but an understandable concern that some exceptionally large loans it had made in the course of the fourth war between the Dutch Republic and England (1780–1784) had gone sour.

¹⁸The notes of New York State "free banks" even announced on their faces that they were "secured by the pledge of public stocks," a clear indication that the notes were backed by something *other* than 100 percent reserves. This inscription was, however, required by law (Hildreth 1968, p. 202).

Fourth, fractional-reserve banking has never been compulsory. Depositors have always been free to insist on 100 percent reserves. They can do so even now, by hiring safety-deposit boxes and stuffing them with cash. (Some do, but mainly to hide their wealth rather than to secure it against bank failure.) Few people have taken the 100-percent-reserve option because—as Rothbard (1990, p. 47) forthrightly acknowledges—it means foregoing interest and paying warehousing fees instead. Most depositors would rather receive interest on their deposits, and consider it more than adequate compensation for the risk involved in fractional-reserve banking. (Here again, we are drawing on evidence from banking systems with relatively unhampered banks and no government deposit guarantees.)

We infer, in accordance with the Rothbardian notion of “demonstrated preference” (Rothbard 1957), that the vast majority of consumers have preferred fractional-reserve banking. Against this Hoppe offers his *a priori* conviction that most depositors could not, would not, and did not ever knowingly engage in such a risk-return tradeoff. For Hoppe the offer of interest on fractionally backed demand deposits is just a swindler’s come-on, which millions of depositors have unwittingly fallen for, wholly innocent of the fact that banks can generate the revenues that go to pay the interest only by lending out some fraction of their deposits.

The fact that banks compete for depositors poses a problem for Hoppe’s position that cannot be so casually brushed aside. Rivalrous competition by fractional-reserve banks seeking depositors’ funds will bid up deposit interest rates (and increase the level of services provided) to the point where banks have to pay such high interest (and provide so many services) to attract deposits that entry is no longer attractive. Thus competition will beat down the returns to capital invested in fractional-reserve banking until the marginal bank is earning only the normal rate of return. In this situation, were it really true that most depositors are willing to forego the interest they are receiving (and instead pay storage fees) in order to have the security of a 100-percent-reserve bank—but simply don’t realize that their banks aren’t holding 100 percent reserves—then any banker (who *does* know what the banks are up to, after all), possessing even an ounce of entrepreneurial insight, would see an easy way to grasp pure profit. All the banker has to do is to offer credible 100-percent-reserve accounts, while alerting the public to the other bankers’ practices, and depositors will come flocking in.¹⁹ If

¹⁹Picture a television spot showing a gleaming vault filled with cash. An authoritative voice-over announces: “Here at the Solid Gold Warehouse Bank, your deposit is

100-percent-reserve banks are legal and really would be preferred by the majority of informed depositors, and the only reason depositors continue to patronize existing banks is ignorance of their fractional practices, then there would be a huge profit to be had by being the first to inform depositors and to offer them the alternative practice.

There have been historical banking systems where explicit 100-percent-reserve banks could have entered the market and where deposit insurance did not exist to slant the playing field in favor of fractional-reserve banks. Yet very few (if any) banks, after the earliest days of banking, have ever tried to attract depositors on that basis. Even if there were one or two such banks in the early days, clearly their approach never spread to dominate the banking market the way it would have if most depositors were truly ready to pay the fees necessary for 100-percent-reserve banking. Maybe entrepreneurship doesn't tend to sniff out profits as well as the Austrian theory of the market process usually suggests. We think it more likely that 100-percent-reserve banking is just not very widely demanded, because of its foregone-interest cost.²⁰

The Resource Cost Savings From Fiduciary Media

Hoppe (pp. 56–58) considers but rejects a standard economic argument we accept concerning fractional-reserve banking: that it reduces the resource costs associated with indirect exchange, by partially substituting bank-issued exchange media for commodity money, thereby reducing (inframarginally) the resource costs of producing money. The resource-cost-saving view is expounded not only by Adam Smith but also by Ludwig von Mises. In *The Theory of Money and Credit*, Mises (1980, p. 333) observes that, thanks to the development of fiduciary media and clearing systems among their issuers, a “tremendous increase in the exchange value of money, which otherwise would have occurred . . . has been avoided, together with its undesirable consequences.” The “undesirable consequences” are the diversion of capital and labor “from other branches of production to the production of the monetary metal.” Had it not been for the development of fiduciary media, Mises points out, “the welfare of the community would have suffered” because “a smaller quantity of economic goods would have been available for the direct satisfaction of human wants.”²¹

backed with genuine 100 percent reserves. *All* your money stays here waiting for you *all the time*. We're not like those *other* banks [camera pulls back to show an adjacent vault which is empty, with moths flying about inside] that try to get by on (gasp!) *fractional reserves*.”

²⁰It might be said that most people would rather “put their money where the moths are.”

²¹For an extended secondary account of Mises's defenses of fractional-reserve banking, see White (1992).

We are puzzled that Rothbard (1990, pp. 33–34), while emphasizing the point that once an economy is fully monetized there is no benefit to money-users from producing more units of money, does not follow Mises in recognizing the consequent value of economizing on the resources used to produce more money.²²

Although many mainstream economists believe that a fiat base money is less costly than a commodity base money, we do not share that view. Fiat money is different because its introduction is involuntary, so that it does not pass a demonstrated preference test, and because its quantity is subject to arbitrary expansion by its issuer, making a fiat system potentially very costly for the economy even if the monetary demand for gold and thus the costs of gold mining were reduced.²³ Our position is rather that, *given* a commodity standard, informed money-users benefit when those who want to are allowed to hold fractionally backed notes and demand deposits. Potential gains from voluntary trade are lost when the public is restricted to full-bodied coins and 100-percent-reserve deposits.

Hoppe (p. 58) denies that redeemable bank monies can save resources. The savings are illusory, he thinks, because “the overwhelming bulk of the population would employ money proper for most of their purchase or sales.” In a footnote (p. 58 n. 11) he adds, without citing a source of evidence: “Indeed, historically this has been the case: Traditionally, notes have always been widely distrusted, and their acceptability—as compared to that of genuine money such as gold or silver coins—was severely limited.”

The facts are otherwise. Throughout the silver and gold standard eras, consumers given a choice ordinarily demonstrated a marked preference for banknotes over full-bodied coins as a more convenient medium of exchange for all but the smallest transactions. The demonstration of preference was especially clear where banking was relatively unhampered by legal restrictions. In Scotland during the free banking era (1716–1844), according to Checkland (1975, p. 382), the first object of any recipient of a gold sovereign was “to get quit of it in exchange for a bank note.” Virtually all sizable payments were made with banknotes.

²²Rothbard (1990, p. 34) argues that gold mining is not socially wasteful, even though an increased supply of gold does not confer any monetary benefit, because gold is a useful commodity for making jewelry, filling teeth, and so on. But the question of social waste from imposing a binding 100-percent-gold-reserve requirement on banks does not concern the cost of mining gold for *non-monetary* uses. It concerns the cost of mining that *portion* of the gold supply destined for bank vaults, over and above the amount of gold banks would acquire if free to choose their own reserve ratios.

²³In practice, the relative price of gold has risen since the scuttling of the gold standard, because few central banks have sold off their gold reserves and because the public has understandably accumulated gold as an inflation hedge. See Garrison (1985).

Similar practices prevailed in Canada (National Monetary Commission 1910, p. 53).

Inherent Instability

Apart from the “fraud” issue and third-party wealth effects, Hoppe believes that fractional-reserve banking is a bad thing because it supposedly produces a monetary instability that contributes to credit cycles and banking crises. We share the view that monetary instability contributes to cycles and crises, but we attribute monetary instability to central banking and other government intervention in the monetary system, not to fractional reserves *per se* or to the practices of competing fractional-reserve commercial banks.²⁴

Hoppe views fractional-reserve banking as something that a proper legal code would ban, and instability as a problem inherent in fractional-reserve banking, and therefore does not distinguish the effects of free banking from the effects of government intervention. Nor does he offer any historical evidence that might test his view against our view. He does take issue with our theoretical argument that free banking tends to permit expansion of the stock of fiduciary media only to an extent consistent with the preservation of monetary equilibrium and the avoidance of the credit-expansion-induced business cycle.

In discussing the requirements for preserving “monetary equilibrium” (that is, equality between the nominal quantities supplied and demanded of money balances, or equivalently between the real stock and real quantity demanded) it is important to distinguish between short-run and long-run implications of changes in the demand schedule for money or in the stock of money. In the long run, nominal prices will adjust to equate supply and demand for money balances, whatever the nominal quantity of money.²⁵ It does not follow, however, that each and every *change* in the supply of or demand for money will lead *at once* to a new long-run equilibrium, because the required price adjustments

²⁴Our writings on cycles and crises include Selgin (1989; 1992; 1993) and White (1984, pp. 18–19, 44–9, 53, 103–12; 1993). Hoppe’s claim that White “nowhere even mentions the problem of business cycles” is easily shown to be false. Even a cursory glance at the index of the only work of White’s that Hoppe cites reveals several mentions of the problem of business cycles (White 1989, pp. 6, 77, 81–4, 142, 159). White (1992, esp. pp. 524–26 and 532, n. 29) directly addresses Mises’s view of banking and the business cycle, including the “Austrian–Misesian claim that *any* injection of fiduciary credit must result in a boom-bust cycle” that is the jumping-off point for Hoppe’s economic critique of free banking. It should be noted that Mises did *not* share Hayek’s view (see White 1995) that fractional-reserve commercial banks, unprompted by central bank policy, can be expected to over-expand and thereby to generate business cycles repeatedly.

²⁵Hoppe (p. 65 n. 19) appropriately refers to this as an “old—Humean—insight.”

take time. They take time because not all agents are instantly and perfectly aware of changes in the money stock or money demand, and because some prices are costly to adjust and therefore “sticky.” It follows that, in the short run (empirically, think “for a number of months”), less than fully anticipated changes to the supply of or demand for money can give rise to monetary disequilibrium. The quantity of money supplied may exceed the quantity demanded, in which case prices need generally to rise; or the quantity of money demand may exceed the quantity supplied, in which case prices need to fall (Yeager 1986).

Such states of monetary disequilibrium, although temporary, may involve serious misallocations of resources. In addition to involving prices that are *generally* “too low” or “too high” (for equilibrium in money holding), they also typically involve distortions of *relative* prices, most importantly (we learn from the Austrian business cycle theory) the rate of interest. Following Wicksell, the Austrian theory holds that an unanticipated injection of money (or rise in the “velocity” of money) can drive the interest rate in the short run below its equilibrium (“natural”) level, and thereby encourage unwarranted investments. Correspondingly, an unanticipated destruction of money (or drop in “velocity”) can drive the interest rate in the short run above its natural level, and thereby artificially curtail warranted investments.

Some economists deny the importance or even the conceptual coherence of short-run monetary disequilibrium as sketched above. New-Classical theorists do so, with a certain internal consistency, because they subscribe to a Walrasian model implying instantaneous and complete price adjustment. Some Austrians do so, with a regrettable *inconsistency*, when they recognize the destructive consequences of price inflation driven by monetary expansion, but nonetheless try to argue that price *deflation* is always okay, in any amount. It is inconsistent to apply short-run, Wicksellian, disequilibrium analysis when talking about increases in the stock of money and price inflation, and then switch exclusively to a long-run, Humean, equilibrium-always analysis when talking about increases in money demand and deflation.

We aspire to be consistent Wicksellians, and so regard both price inflation and deflation as regrettable processes *insofar as they are brought about by arbitrary changes in the nominal quantity of money, or by uncompensated changes in its velocity, and not by changes in the real availability of final goods or the cost of production of money* (Selgin 1990, 1995; White 1990). It is therefore an attractive feature of free banking with fractional reserves that the nominal quantity of bank-issued money tends to adjust so as to offset changes in the velocity of money (Selgin and White 1994, p. 1725). Free banking thus works

against short-run monetary disequilibrium and its business cycle consequences.

The argument for the equilibrating properties of free banking rests in part on recognizing that an increased demand to hold claims on intermediaries, including claims in the form of banknotes and demand deposits, at the expense of holding additional consumer goods, is equivalent to an increase in desired saving. Hoppe (p. 72) disagrees, labeling this analysis a “confusion.” He declares that

it is plainly false to say that the holding of money, i.e., the act of not spending it, is equivalent to saving. One might as well say—and this would be equally wrong—that the not-spending of money is equivalent to *not*-saving. In fact, saving is not-consuming, and the demand for money has nothing to do with saving *or* not-saving.

We submit that the confusion is Hoppe’s, not ours. The above-quoted passage identifies saving as not-consuming, which taken literally means that saving is any disposition of wealth other than for present consumption. Elsewhere (p. 50) Hoppe correctly observes that money “is demanded as a medium of exchange—rather than for consumption or production purposes,” that is, that money-holding is a form of not-consuming. Together these statements contradict his claim that holding money is not a form of saving.

Hoppe’s position is that saving is an expression of time preference, but money-holding is not. Thus to save is to *defer* consumption, and because the holding of money does not signal a definite decision to defer consumption (unlike the purchase of a bond or a capital good), it is not a form of saving.²⁶ We agree that time preference and money demand are distinct, and that a change in one does not imply a change in the other. Nonetheless, to hold money *is* to hold it for *later* spending, even though *how much later* is not signalled (and typically has not yet been decided by the money-holder). Holding money for later spending, rather than spending it on consumption *now*, does defer consumption to the future. As Hoppe (pp. 72–3) himself points out, the demand for cash stems from the convenience it allows one in purchasing “consumer or producer goods at uncertain *future* dates” (emphasis added). So perhaps our disagreement here is merely over words.

The substantive question Hoppe raises is whether, as he asserts, “any injection of fiduciary credit must result in a boom-bust cycle.” We deny that an increase in fiduciary media *matched by an increased demand to hold fiduciary media* is disequilibrating or sets in motion the

²⁶Thus Hoppe (p. 72) emphasizes that to hold money “is to purchase neither consumer goods *nor* investment goods.”

Austrian business cycle. The act of holding fractional-reserve *bank-issued* money not only (like holding base money) defers consumption for a longer or shorter period, but also *temporarily lends funds* to the bank of issue in so doing. The period of the loan is unspecified—a demand deposit or banknote can be redeemed at any time, though only a fraction are in fact redeemed on any day—but if the bank can estimate with a fair degree of accuracy the lengths of time for which its demand claims will remain in circulation (the statistical distribution of their times to actual redemption), it can safely make investments of corresponding length.²⁷ As Mises (1980, p. 362) wrote with respect to the related problem of estimating the volume of demand for a bank's fiduciary media, the banker here “has to rely upon an uncertain empirical procedure which may easily lead to mistakes. Nevertheless, prudent and experienced bank directors—and most bank directors are prudent and experienced—usually manage pretty well with it.”

De Soto (1995, p. 32) asserts that fractional-reserve free banking “must inexorably, sooner or later, lead to uncontrolled expansion in the monetary supply,” and claims Mises's authority for this view. Mises (1966, p. 443) actually, and we believe quite correctly, held a very different view:

Free banking is the only method for the prevention of the dangers inherent in credit expansion. It would, it is true, not hinder a slow credit expansion, kept within very narrow limits, on the part of cautious banks which provide the public with all the information required about their financial status. But under free banking it would have been impossible for credit expansion with all its inevitable consequences to have developed into a regular—one is tempted to say normal—feature of the economic system. Only free banking would have rendered the market economy secure against crises and depressions.

Hoppe misunderstands Selgin's argument when he characterizes it as jumping from the view that the holding of money represents savings to the conclusion that “an increased demand for money [is] the same thing as increased saving.” That holding money is one form of saving does not imply that an increase in the demand for money is identically an increase in total saving. An increased demand for money may accompany a reduced demand for holding other assets, and not a reduction in consumption; hence it may be part of a change in the manner of saving with no change in total savings. If, for example, the public's demand for bank deposits increases at the expense of the public's demand for bonds, holding

²⁷Thus interest-bearing demand deposits are not inconsistent with sound banking.

constant the rate of time preference (or, alternatively put, holding constant the planned and expected time-profile of consumption),²⁸ there will be no change in “the” natural rate of interest, viewed as a composite of interest rates on all financial assets. Expansion of the volume of deposits is nonetheless warranted in this case. Assuming rising marginal costs of intermediation, the equilibrium rate of interest on bank deposits will have fallen, while the rate on bonds will have increased. The increased demand for intermediation raises the “price of intermediation” represented by the spread between the deposit and bond rates. Banks are warranted in expanding their balance sheets to meet the increased demand for deposits, until the actual deposit rate falls to the new equilibrium deposit rate. (Meanwhile, the market value of existing bonds falls *pari passu* with the increase in the bond interest rate.)

An increase in *savings* is neither necessary nor sufficient to warrant an increase in fiduciary media. An increased demand for “cash” (Hoppe, p. 73) does not warrant an increase in fiduciary media or inside money, assuming that “cash” is used to mean *high-powered* or *outside* money such as gold coins (as opposed to low-powered, competitively-issued banknotes). It is specifically an increased demand to hold “balances of inside money” (Selgin’s words, quoted by Hoppe) that warrants an increase in the quantity of inside (bank-issued) money. A banking system that accommodates an increased real demand to hold its demand liabilities by expanding their quantity does nothing to drive market interest rates away from their natural values, spur excessive investment, or set in motion a boom-bust cycle.

We can put this point another way. Consider the case in which the public increases its desire to save, due to a drop in time preference, and people elect to forego some current consumption spending out of their income in order to build up their holdings of *time* deposits issued by banks. We imagine that no Austrian will object that it is dangerous to allow the banking system to accommodate this shift. The natural rate of interest has fallen. The public correspondingly bids down the interest rate on time deposits, and by lending their extra deposits banks bid down the interest rate on loans, so the market rates correctly track the natural rates.

Now consider the case where the public increases its desire to save, due to a drop in time preference, and people elect to forego some current consumption spending out of their income in order to build up

²⁸It may be that this *ceteris paribus* condition is seldom met in practice. It may be that a shift from bonds to money is usually joined to a change in time preference, i.e., is usually accompanied by a shift toward the present (or toward the future) in the planned time-profile of consumption. Nonetheless these shifts are conceptually distinct. The *ceteris paribus* assumption allows us to analyze their effects separately.

their holdings of interest-bearing *demand* deposits issued by banks. We submit that it is no more dangerous, or disequilibrating, or cycle-inducing, to allow the banking system to accommodate *this* shift.²⁹ It would, instead, be disequilibrating and unfortunate if the banking system were *not* to respond. The velocity of bank-issued money (the ratio of dollars spent per year to dollars held) has fallen. If the banking system fails to increase the quantity of bank-issued money and the price level does not promptly drop, an excess demand for money arises (assuming also that the quantity of base money does not immediately increase). A corresponding excess supply of goods arises: unsold consumer goods pile up on sellers' shelves (this is of course what proximately puts downward pressure on prices, until at last goods prices have fallen sufficiently). Business is depressed until the purchasing power of money gets back to equilibrium. By failing to increase the quantity of deposits, the banking system also fails to bid down the interest rate on deposits and loans. The natural rate of interest has fallen, but market interest rates temporarily stay put. Investment does not increase to match the increased desire to save, and the structure of production does not adapt as it should to match the drop in time preference.

Conclusion

Fiduciary media are not fiat money. A monetary system with a commodity standard, competitive banking, and the freedom to use fiduciary media among consenting transactors is consistent with justice, efficiency, and economic stability. It is preferable on these scores both to a system (like today's) where the law has forced money-users to give up gold and gold-redeemable fiduciary media in favor of fiat money, and to a system (like those proposed by 100-percent-reserve advocates) where the law restricts money-users from holding any or some types of fiduciary media.

²⁹But how can the banks manage to expand their demand deposits, if total bank reserves have not changed? The increased demand to *hold* demand deposits, relative to income, means that fewer checks are written per year per dollar of account balances. The marginal deposit dollar poses less of a threat to a bank's reserves. Thus a bank can safely increase its ratio of deposits to reserves, increasing the volume of its deposits to the point where the rising liquidity cost plus interest and other costs of the last dollar of deposits again equals the marginal revenue from a dollar of assets (Selgin 1994b).

References

- Block, Walter. 1988. "Fractional Reserve Banking: An Interdisciplinary Perspective." In Walter Block and Llewellyn H. Rockwell, Jr., eds., *Man, Economy, and Liberty: Essays in Honor of Murray N. Rothbard*. Auburn, Ala.: Ludwig von Mises Institute.
- Checkland, S. G. 1975. *Scottish Banking: A History, 1695–1973*. Glasgow: Collins.
- de Soto, Jesús Huerta. 1995. "A Critical Analysis of Central Banks and Fractional-Reserve Free Banking from the Austrian Perspective." *Review of Austrian Economics* 8 (2): 25–38.
- Dowd, Kevin, ed. 1992. *The Experience of Free Banking*. London: Routledge.
- Garrison, Roger W. 1985. "The Cost of a Gold Standard." In Llewellyn H. Rockwell, Jr., ed., *The Gold Standard: An Austrian Perspective*. Lexington, Mass.: Lexington Books.
- Gorton, Gary B. 1988. "Banking Panics and Business Cycles." *Oxford Economic Papers* 40 (December): 751–81.
- Hayek, F. A. 1978. *The Denationalisation of Money*, 2nd ed. London: Institute of Economic Affairs.
- Hildreth, Richard. 1968. *Banks, Banking, and Paper Currencies*. New York: Greenwood Press.
- Hoppe, Hans-Hermann. 1994. "How is Fiat Money Possible?—or, the Devolution of Money and Credit." *Review of Austrian Economics* 7 (2): 49–74.
- Hülsmann, Jörg Guido. 1995. "Free Banking and the Free Bankers." *Review of Austrian Economics* 9 (1): 3–53.
- Kaufman, George. 1994. "Bank Contagion: A Review of the Theory and Evidence." *Journal of Financial Services Research* 8 (April): 123–50.
- Mises, Ludwig von. 1966. *Human Action: A Treatise on Economics*. 3rd rev. ed. Chicago: Henry Regnery.
- . 1980. *The Theory of Money and Credit*. H. E. Batson, trans. Indianapolis: Liberty Classics.
- National Monetary Commission. 1910. *Interviews on the Banking and Currency Systems of Canada*. Washington: Government Printing Office.
- Rothbard, Murray N. 1957. "Toward a Reconstruction of Utility and Welfare Economics." In Mary Sennholz, ed., *On Freedom and Free Enterprise*. Princeton: Van Nostrand.
- . 1962. "The Case for a 100 Percent Gold Dollar." In Leland B. Yeager, ed., *In Search of a Monetary Constitution*. Cambridge: Harvard University Press.
- . 1970a. *Man, Economy, and State: A Treatise on Economic Principles*. Los Angeles: Nash.
- . 1970b. *Power and Market: Government and the Economy*. Menlo Park, Calif.: Institute for Humane Studies.
- . 1983a. *The Ethics of Liberty*. Atlantic Highlands, NJ: Humanities Press.
- . 1983b. *The Mystery of Banking*. New York: Richardson & Snyder.
- . 1990. *What Has Government Done to Our Money?* 4th ed. Auburn, Ala.: Ludwig von Mises Institute.
- . 1995. "Fractional Reserve Banking." *The Freeman* 45 (October): 624–27.

- Salsman, Richard M. 1990. *Breaking the Banks: Central Banking Problems and Free Banking Solutions*. Great Barrington, Mass.: American Institute for Economic Research.
- Selgin, George A. 1989. "Legal Restrictions, Financial Weakening, and the Lender of Last Resort." *Cato Journal* 9 (Fall): 429–59.
- . 1990. "Monetary Equilibrium and the 'Productivity Norm' of Price-Level Policy." In Richard M. Ebeling, ed., *Austrian Economics: Perspectives on the Past and Prospects for the Future*. Hillsdale, Mich.: Hillsdale College Press.
- . 1992. "Bank Lending 'Manias' in Theory and History." *Journal of Financial Services Research* 6 (March): 169–86.
- . 1993. "In Defense of Bank Suspension." *Journal of Financial Services Research* 7 (December): 347–64.
- . 1994a. "Are Banking Crises Free-Market Phenomena?" *Critical Review* 8 (Fall): 591–607.
- . 1994b. "Free Banking and Monetary Control." *The Economic Journal* 104 (November): 1449–59.
- . 1994c. "On Ensuring the Acceptability of a New Fiat Money." *Journal of Money, Credit, and Banking* 26 (November): 808–26.
- . 1995. "The 'Productivity Norm' vs. Zero Inflation in the History of Economic Thought." *History of Political Economy* 27 (Winter): 705–35.
- Selgin, George A., and Lawrence H. White. 1987. "The Evolution of a Free Banking System." *Economic Inquiry* 25 (July): 439–57.
- . 1994. "How Would the Invisible Hand Handle Money?" *Journal of Economic Literature* 32 (December): 1718–49.
- Smith, Adam. 1981. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Indianapolis: Liberty Classics.
- Van Dillen, J. G. 1934. "The Bank of Amsterdam." In Van Dillen, *History of the Principal Public Banks*. The Hague: Martinus Nijhoff. Pp. 79–123.
- White, Lawrence H. 1984. *Free Banking in Britain*. Cambridge: Cambridge University Press.
- . 1989. *Competition and Currency: Essays on Free Banking and Money*. New York: New York University Press.
- . 1990. "Commentary: Norms for Monetary Policy." In Richard M. Ebeling, ed., *Austrian Economics: Perspectives on the Past and Prospects for the Future*. Hillsdale, Mich.: Hillsdale College Press.
- . 1992. "Mises on Free Banking and Fractional Reserves." In John W. Robbins and Mark Spangler, eds., *A Man of Principle: Essays in Honor of Hans F. Sennholz*. Grove City, Penn.: Grove City College Press.
- . 1993. "Why is the U. S. Banking Industry in Trouble? Business Cycles, Loan Losses, and Deposit Insurance." In Lawrence H. White, ed., *The Crisis in American Banking*. New York: New York University Press.
- . 1995. "The Evolution of Hayek's Monetary Economics." Unpublished ms. University of Georgia.
- Yeager, Leland B. 1986. "The Significance of Monetary Disequilibrium." *The Cato Journal* 6 (Fall): 369–99.

Central Banking, Free Banking, and Financial Crises

Roger W. Garrison

A growing literature explores the concept of free banking on both a theoretical and an historical basis. George Selgin (1988) sets out the theory of free banking and makes a compelling case that, despite the uniqueness of money, the forces of supply and demand are more conducive to monetary stability, correctly understood, than are the edicts of a central bank. Larry White (1984), focusing on the free-banking episode in nineteenth-century Scotland, and Kevin Dowd (1994), collecting studies of experience with free banking in many countries and time periods, have shown that this alternative to central banking has a respectable history.

The aim of this paper is to get a fix on the possible and currently relevant sources of macroeconomic instabilities in the economy and to identify the most promising banking arrangements for dealing with those instabilities. Possible maladies and remedies can be considered in the context of competing schools of macroeconomic and monetary thought. Attention is directed to the issue of whether the perceived problem and/or its solution is inherent in the market economy or lies outside the market process. This formulation immediately gives rise to a two-by-two matrix with maladies and remedies represented in one dimension, market forces and extramarket forces represented in the other. The fruitfulness of this approach is demonstrated by its ability to sort out competing schools of thought, put current debate in perspective, and assess the prospects for a stable macroeconomy—with the Federal Reserve as currently constituted and with the alternative institution of free banking.

This exercise in comparative-institutions analysis does not deal with the dynamics of the macroeconomy in transition between one set of monetary institutions and another or with the political issues of just how such a transition might be brought about. Nor does it deal directly

*Roger Garrison is professor of economics at Auburn University.

The Review of Austrian Economics Vol. 9, No. 2 (1996): 109–27
ISSN 0889-3047

with the ultimate nature of the monetary standard. There is a strong presumption, however, that only a central bank can preempt a commodity standard with its own fiat money and that banknotes issued by competing banks in a free-banking system would have to be redeemable in some real commodity, such as gold, to make them acceptable in a market where banknote holders can easily express their preferences among issuers. There is broad agreement among Austrian-oriented writers that a banking system characterized by (1) central direction and (2) fractional reserve is not conducive to economic stability. However, there is some disagreement among the Austrians as to which of the two mentioned characteristics is fundamentally responsible for the instability. The argument in this paper follows Ludwig von Mises, as portrayed by White (1992), and takes the centralization of the current banking system to be the most fundamental issue and the most appropriate focus for prescribing reform.

The Equation of Exchange

Underlying all theories of money and banking—as well as all prescriptions of policy and recommendations for reform—is the familiar equation of exchange: $MV = PQ$. For the economy as a whole, buying must equal selling, where buying is represented by the total supply M of money times the frequency (the circulation velocity V) with which each monetary unit on average is spent and where selling is represented by the average price P of goods times the total quantity Q of goods sold. Although true by construction, the equation of exchange helps us to keep in view the interdependencies that characterize the macroeconomy. It is impossible, for instance, to conceive of a change in only one of the four magnitudes represented in the equation of exchange. Any one change implies some offsetting change or changes on one side or the other of the equation—or possibly on both sides. For instance, a decrease in money's circulation velocity, which simply reflects an increase in the demand for money, must be accompanied by (1) an increase in the money supply, (2) a decrease in prices, or (3) a decrease in real output sold (or by some combination thereof).

The equation also facilitates the comparison of competing schools of thought. Considering in sequence Keynesianism and Early and Late Monetarism can provide a basis for setting out the distinctive perspective that emerges from the theory of free banking.¹ The case

¹The comparison of schools facilitated by the equation of exchange is wholly independent of the unique qualities of Austrian macroeconomics, which features the intertemporal allocation (and possible misallocation) of resources and requires theorizing at a lower level of aggregation.

against central banking and in favor of free banking, then, is preceded by some history of thought—possibly more than some may think justified. The comparison of schools of thought is included for two reasons. First, some writers have recently gotten it wrong, presenting monetarist ideas under the Keynesian label. Second, the case for free banking contains arguments that are sufficiently close to Keynes's own that they need to be distinguished explicitly from his.

Keynes believed that the economy is chronically unstable because of instabilities associated with both Q and V . Goods, in the Keynesian construction, are decomposed into consumption goods C and investment goods I , the latter being inherently unstable in view of the pervasive uncertainty faced by the business community—the “dark forces of time and ignorance that envelop our future” (Keynes, 1936, p. 155). The strength of the investment sector, according to Keynes, is highly dependent on psychological factors—“animal spirits” (pp. 161–62) that motivate each (and, through contagion, all) of the economy's investors. The occasional waxing and waning of the animal spirits affect I —and affect C as decisions in the business community govern incomes and hence spending. Both directly and derivatively, then, the uncertainty of the future translates into fluctuations in the economy's output magnitude Q .

The equation of exchange reminds us that changes in Q cannot be the whole story. If prices and wages are sticky and the money supply is wholly determined by the monetary authority, the rest of the story must center on money's circulation velocity V . What Keynes called the “fetish of liquidity” is, in this view, nothing but another perspective on the waning of “animal spirits.” Would-be investors abstain from committing themselves to investment projects, whose profitability is uncertain, and instead hold their wealth liquid.

The economy, according to Keynes, is prone to periodic collapse. Pervasive uncertainty inherent in investment activity and prospects of economic disaster occasionally overwhelm the business community. Entrepreneurs cease their individual attempts to outguess one another and begin collectively to guess against the economy. In droves, they forego real assets in favor of liquidity. Q falls, and along with it, V . Liquidity, or money (Keynes used the terms synonymously), constitutes something of a “time out” for the entrepreneur/speculator—somewhat analogous to rest areas along an interstate highway. Fog on the highway or the wearing effects of traffic congestion can make the rest areas increasingly attractive.

The origin and essence of the problem, in the Keynesian view, is to be found on the righthand side of the equation of exchange (a decreased Q). Keynes works on both sides of the equation, however, in

devising possible solutions to the problem. For instance, much of Keynes's discussion of monetary reform, which included support in principle for Silvio Gesell's stamped money as well as for taxing transactions in securities markets, was aimed at making the time-out option—the option of getting or staying liquid—more costly. Keynes favored all attempts to deprive money of its liquidity value only to lament that investors would find other assets (e.g., gems and precious metals) that could provide refuge from the uncertain future (Keynes 1936, pp. 353–58).

Reforms in this direction are analogous to installing toll gates at the rest areas—or possibly eliminating rest areas altogether. Travelers would make better time between New Orleans and Atlanta if there were no possibility of stopping along the way. Keynes did not consider that some would-be travelers might not depart New Orleans in the direction of Atlanta under such conditions; he did lament that closing or charging for rest areas might cause travelers to find other places to stop along the highway.

In lieu of prevention in the form of making liquidity less attractive or more costly, Keynes recommended monetary policy to accommodate the demand for liquidity—satiating that demand if necessary to keep money from competing with real investments in the collective mind of the business community. To the extent that money-demand entails a large psychological element, the rest-area analogy holds. A road sign that reads “LAST REST AREA FOR NEXT 100 MILES” may attract many customers, whereas the travelers may stop very infrequently if there were rest areas all along the way.

While increasing the supply of money to neutralize the effects of a fetishistic demand for liquidity may be a necessary component of policy prescription, it will not be sufficient, according to Keynes, to restore conditions of prosperity. This is only to say that a decreased V is a symptom rather than the essence of the problem. The solution must involve the substitution of government spending for private investment spending—accommodated, of course, by money creation. Fiscal stimulation prods the reluctant travelers along the economic highway. Keynes viewed fiscal policy as primary; monetary policy as secondary.

In the Keynesian view, then, the malady is inherent in the market; the remedy entails extramarket forces. It is in the very nature of things that our weary travelers will, on occasion, follow one another into the increasingly overcrowded rest areas, where each traveler is reluctant to resume the journey alone. Restoring and maintaining stability requires intervening forces in a double-barreled way; the interveners must work simultaneously on both sides of the equation of exchange.

Monetary reform and fiscal stimulation are intended to keep the travelers out of the rest areas and to keep them moving along smartly. Central banking is essential for the task. But ultimately, Keynes (1936, p. 378) called for a wholesale replacement of our current system with a system of public transportation: A comprehensive socialization of investment is offered as the only solution to the problem of unemployment.

Early monetarism, as expounded by Clark Warburton (1966) in the 1940s and 1950s and as revived in recent years by Leland Yeager (1986), has a kinship to the equation-of-exchange perspective on the Keynesian view. Both schools perceive a possible malady and remedy that fit into the two-by-two matrix in the same way: Market malady; extramarket remedy. They differ radically, however, in terms of the specific nature of the problem and the implied judgment about the efficacy of the market economy. Market participants may opt for more money in preference to more real output—where the relevant alternatives to holding money are both investment goods and consumption goods. The demand for money is not fetishistic, and changes in it are not necessarily contagious, but money demand can and does change. The velocity of money is not constant in the same way that Planck's constant and Avogadro's number are.²

With a given money supply, increases in the demand for money put downward pressure on prices.³ Except in the fanciful case in which prices adjust fully and instantaneously to this monetary disturbance, the adjustment process involves quantities as well as prices. Our highway travelers are trying to stop and rest even in the absence of adequate rest areas. The unintended consequence is a general slowdown of traffic. A decreased V impinges on Q as well as on P —even if the ultimate, or long-run consequence is a proportionate decrease in P . In principle, a monetary policy that succeeds in relieving downward pressure on prices by meeting every increased demand for money with an increased supply will result in greater stability for the economy as a whole. A constant P becomes, in this view, the essence of monetary stability. The problem (decreased V) and solution (increase M) are set out in precisely this way by Paul Krugman (1993, p. 26–28 and *passim*)—but

²It should be noted, however, that even before the impact of Milton Friedman's empirical work was fully felt, the Early Monetarists held that the typical and most significant reductions in MV were attributable to reductions in M and not in V .

³Here and throughout the paper, the phrases "increase in the demand for money" and "decrease in the velocity of money" are used interchangeably. Although this usage is not unconventional, some monetary theorists take money demand to be defined by the equation of exchange itself. That is, $M_d = (1/V)PQ$, in which case *any* change on the righthand side of the money-demand equation would constitute a change in the demand for money.

with this view offered as Keynes's understanding of the nature of business cycles! Early Monetarism is wrongly attributed to Keynes.⁴

Early and Late Monetarists share an analytical framework as well as a basic judgment about the central bank's capacity to do good and to do harm. It was Milton Friedman, of course, who shifted the focus of attention away from problems of monetary disequilibrium to the general relationship between M and P that endures over space and time. Empirical studies using data from many different economies and many different time periods lent support to the proposition that changes in the lefthand side of the equation of exchange are overwhelmingly attributable to changes in the quantity of money. Study after study demonstrating the stability of money demand (a near-constant V) had the effect of focusing attention on the money supply M as a basis for accounting for both inflation and deflation. Changes in the money supply are much more likely to be a problem than to be a solution to a problem. Empirical and theoretical considerations, as well as considerations from political economy, underlay this summary judgment. Under typical conditions, in which money demand remains relatively constant, there is a "long and variable lag" that separates changes in the money supply and the subsequent changes in the price level. This empirical fact, coupled with the lack of any timely and unambiguous indicator of actual changes in the demand for money, weighs against the prospects for even well-intentioned money-supply management having a stabilizing effect on the macroeconomy. Dimming the prospects still further, of course, is the fact that the central bank may intend to do more than act as a stabilizing agent and that some of its intentions, such as dealing narrowly in alternating episodes with the problems of inflation and unemployment and with problems associated with the strength or weakness of the dollar in international markets, are antithetical to the idea of a central bank as macroeconomic stabilizer.

⁴Even worse, the school of thought whose sails have most recently caught the academic wind calls itself New Keynesianism—seriously missing the mark with both parts of its name. Gregory Mankiw and others (Ball, et al., 1988) remain largely agnostic about the specific source of change on the lefthand side of the equation of exchange. Their theorizing holds up whether it is M or V that decreases. The Keynesian label is adopted simply on the basis of their recognition that prices do not change instantly—a basis that actually distinguishes their (and many other) arguments only from extreme versions of New Classicism. The "New" is added in recognition that the assumption of sticky prices is replaced with "sophisticated" reasons for prices not adjusting instantaneously. But Early Monetarism as initially set out and in modern expositions does not fail to include reasons for the behavior of those who set prices. New Keynesianism is Early Monetarism offered with the aid of now fashionable modeling techniques, which involve mathematically tractable—if largely implausible—constraints on price- and wage-adjustments.

We can locate Monetarism in our two-by-two matrix by noting that both malady and remedy are in the extramarket category. In fact, Monetarism consists, by and large, of (1) the recognition that the central bank is a destabilizing force and (2) the recommendation that it *not* be a destabilizing force. Adherence to a monetary rule according to which the money supply is increased at a slow, steady, and preannounced rate is likely to engender more macroeconomic stability than central bank activism can achieve—no matter how well-intentioned and expertly conceived. Actual experience both before and after the heyday of Monetarism suggests that the same understanding that gives rise to Monetarists' view of the central bank also accounts for the central bank's inability and unwillingness actually to adopt and abide by a monetary rule. The so-called Monetarist experiment begun in October of 1979 under the chairmanship of Paul Volcker, for instance, was Monetarist only in a limited and perverse sense. The Federal Reserve did shift its attention from interest rates to monetary aggregates, a move that would be preliminary to actually adopting a rule for monetary growth. But its policies following this shift made for even greater variation in the money supply (and in the rate of interest) creating significantly greater macroeconomic instability than had been experienced before. Ultimately, a monetary rule, however widely and forcefully recommended, is at odds with the even more widely perceived view that the Federal Reserve Chairman is the second most powerful individual in the country.

Free Banking

The basic case for free banking is the general case for decentralization of economic activity. The uniqueness of money does not immunize it against the forces of supply and demand and does not make the invisible hand of the marketplace any less beneficial to society. Quite to the contrary, our rest-area analogy suggests that market forces have special advantages in adjusting money supply to money demand. While the market cannot respond on a daily basis, supplying rest areas anywhere along the highway that they happen to be demanded by today's travelers, free banking can and automatically would supply liquidity along the economic highway anytime and anywhere it is demanded. The case for decentralization is strengthened by comparing free-banking dynamics to central-bank policies that we have actually experienced and even to the policies of an idealized non-politicized central bank whose sole objective is that of maintaining macroeconomic stability. A comparison favoring free banking follows from two propositions. First, the failure in fact of the central bank to adopt a monetary rule (and the unlikelihood of its adopting such a rule in the future)

weighs in favor of decentralization. What the Federal Reserve lacks the will and ability to do can be done automatically by the impersonal forces of supply and demand governing banknote issue. Second, the difference between the implicit rule that the decentralized banking system follows and the simple monetary rule of slow and steady growth of the money supply gives free banking higher marks as a stabilizing force in the economy. In the final analysis, the simplicity of the monetary rule derives from the judgment that discretionary moves are more likely to destabilize than to stabilize. The monetary rule is imposed, then, in the spirit of the unspoken maxim of yesteryear's medical profession: "Maintain good bedside manners, and strive to do no harm."

Free banking automatically discriminates between real disturbances and monetary disturbances, reacting only to the latter (Selgin 1988, pp. 64–69). The "automaticity" implies both a timeliness and an absence of political pressure—features that are forever denied to central banking. Under steady-state conditions in which the economy is experiencing no growth and no changes in the demand for money, the simple monetary rule and the implicit free-banking rule are the same: zero growth in the money supply. The consequences are also the same: a constant price level. Under more typical conditions of some positive rate of real economic growth and some variability in the demand for money, the two rules differ. The simple monetary rule is based on a long-range estimate of secular growth and of secular movement in money demand. An estimated growth rate of 3 percent and an estimated upward trend in money demand (downward trend in velocity) of 2 percent translate into a money growth rate of 5 percent. Strict compliance with the rule would mean that movements in the price level would exhibit no long-run trend. Actual deviations from trend in either output or in velocity, however, would result in upward or downward pressure on the general level of prices. Accordingly, the rule itself might be adjusted to allow for the differential harmfulness of inflation and deflation. Ingrained notions that prices and wages are stickier downwards than upwards and that unemployment bites harder into economic prosperity than does inflation may justify—narrow political motives aside—a rule of increasing the money supply at some rate slightly in excess of 5 percent. A mild inflation might be considered cheap insurance against any actual deflation.⁵

⁵By wholly ignoring disordinating consequences of deflationary pressures and factoring in the effect of an anticipated price-level decline on the real value of money holdings, Friedman (1969, pp. 45–47) argues for a theoretically optimal growth rate for *M* that is considerably *lower* (2% instead of 5%) than that implied by secular changes in *Q* and in *V*.

The implicit rule automatically implemented by free banking is the old central-bank maxim (usually observed in the breach): "Print money to hold but not money to spend." If the holders of banknotes issued by a particular bank are willing to hold still more, it is in the interests of the bank to increase its issue. The fact that the bank's customers are holding rather than spending implies the absence of inflationary pressures. In this context, the bank need not even consider whether the increased demand for its own notes is a general increase in the demand for money or an increase in the demand for its banknotes relative to the demand for other banknotes. However, if an individual bank increases its issue even in the absence of any increase in demand to hold its banknotes, then the extra spending of them will soon impinge on the bank's reserves. The sustainable level of note issue is demand-determined. In a decentralized and competitive environment, each individual bank can be expected to forego the short-term gains that overissuing its own banknotes might entail in order to avoid the long-term losses that the market process would inevitably impose.

In contrast to the simple monetary rule, which is devised to accommodate real economic growth by checking deflationary pressures whatever their source, the implicit free-banking rule involves no change in the money supply in response to a change in real output. This difference in the two rules reflects the automatic discrimination, inherent in free banking, between real and monetary disturbances. An increase in the demand for money puts downward pressure on product and factor prices in general. If there were no money-supply response, a general decline in economic activity would follow, since prices and wages could not fully and instantaneously adjust themselves to the new market conditions. Goods in general would go unsold; production would be cut; workers would be laid off. Such quantity effects can be self aggravating, as the Early Monetarists emphasized. With a less-than-perfectly flexible price system, general deflationary pressures can push the economy below its potential during the period in which prices are adjusting to the higher monetary demand. And the fact that some prices and some wages are more flexible than others means that the adjustment period will involve changes in relative prices that reflect no changes in relative scarcities. These are precisely the kinds of problems that are highlighted by modern monetary-disequilibrium theorists, e.g., Yeager (1986), and that are avoided by free banking's responsiveness to increases in money demand.

Suppose, however, that with an unchanging demand for money, the economy experiences economic growth. Despite the implications of the familiar neoclassical growth models, the economy's output does not undergo a general change; there is no disembodied growth that might

be explained in terms of an economywide technology shock. Rather, the outputs of various goods increase as a result of an increased availability of particular resources used in producing them or the discovery of a new technique that converts particular inputs into a particular output more efficiently. Downward pressure on the prices of the particular goods that account for the economy's growth will be felt primarily in the markets for those very goods. Relative prices adjust to reflect the fact that these goods are now more abundant. The market process at work here is the one that gets emphasis in the sophomore-level economics of supply and demand. Perversities that dominate in the context of an increase in money demand get little or no play in the context of economic growth. The increased Q , which simply reflects a positive net change in the sum of all the economy's individual q_s , is accompanied by a decrease in the corresponding p_s . It would be misleading here to evoke the fears of "deflationary pressures." The individual p_s become adjusted to their corresponding q_s on a market-by-market basis. The fact that this new constellation of p_s average to a lower P than before has no special claim on our attention. There is no downward pressure on P over and above the forces of supply and demand that operate separately in the affected markets and reflect the underlying economic realities. There are no perversities inherent in this sort of a relative (and absolute) adjustment.

In terms of the equation of exchange, we can say that free banking adjusts M so as to offset changes in V ; but allows changes in Q to be accommodated by changes in P . Economic growth does involve price deflation in a literal sense (the price level falls as output increases) but does not involve any macroeconomic malady that is commonly associated with the term "deflationary pressures." In effect, by distinguishing between malignant and benign deflation, free banking provides a much stronger check against inflation than that provided by the simple monetary rule.⁶ It would be misleading to classify free banking in terms of malady and remedy because the malady never gets a chance to show itself. Significantly, though, there are no extramarket forces at work here either creating problems or fixing them.

Central Banking and the Debt Bomb

The case for a decentralized banking system, which by and large parallels the case for markets and against central planning agents, is a

⁶Selgin (1991) distinguishes clearly between what I have called malignant and benign deflation. It is interesting to note that free banking, which relieves only the malignant deflationary pressures, may get close to Friedman's theoretical optimum, which assumes those pressures away. (See footnote 5.)

strong one. The central bank cannot outdo free banking or even match its performance as a macroeconomic stabilizer. It lacks the ability to distinguish on a timely basis between movements in V and movements in Q , it lacks the incentives to act in ways that would promote stability, and as a key player in a political environment, it actually responds to incentives in ways that foster instability. None of these characteristics, however, is at odds with our understanding of the origins of the Federal Reserve System—especially as expounded by Rothbard (1994), whose story does not place great emphasis on the lofty goal of macroeconomic stabilization.

It is commonly understood, now, that the Federal Reserve accommodates the Treasury by monetizing the government's debt. That is, it injects credit markets with new money so as to relieve the upward pressure on interest rates that Treasury borrowing would otherwise entail. And with telling exceptions, the Federal Reserve maintains an easy-money policy in the year-and-a-half before each presidential election.⁷ The so-called political business cycles have now become an integral part of the macroeconomic landscape. Further, the Federal Reserve is called upon to deal with other real or perceived problems having little to do with macroeconomic stability. It is expected, for instance, to lower interest rates when the housing market is in a slump and to strengthen or weaken the dollar in response to movements in exchange rates or trade flows. All these attempts to manipulate employment rates, interest rates, and exchange rates interfere with the Federal Reserve's ability to achieve and maintain macroeconomic stability or even to refrain from inducing instability. If the simple monetary rule fares poorly in comparison with the implicit rule of free banking, it fares well in comparison with the actual policies of the Federal Reserve.

These political factors are well recognized by modern Fedwatchers. Less well recognized are the cumulative effects of decades of deficit accommodation and macroeconomic manipulation. With federal

⁷The telling exceptions involve Presidents Ford, Carter, and Bush. In 1976 Ford simply did not play the game. He did not press Federal Reserve Chairman Arthur Burns, who had helped Nixon get re-elected four years earlier. With Ford perceived as a non-starter, Carter boasted that his administration would "hit the ground running," which in terms of monetary policy meant that the expansion was started much too early. By re-election time (1980), the stimulative effects of the monetary expansion had receded into history and inflation was upon us. With equally bad timing, but in the opposite direction, Bush tried to play the game in 1992 but started the expansion too late—after finally realizing that he couldn't ride through the election on his victory in the Persian Gulf. The monetary stimulant was felt during the first few months of the Clinton administration. Starting too late, too early, and not at all, these three incumbent campaigners had one thing in common: They lost.

indebtedness now measured in the trillions of dollars and increasing annually by hundreds of billions, the need for a stabilizing monetary system is all the more important. The debt bomb is not ignored by Wall Street. An explosive ending to this era of fiscal irresponsibility may or may not be in the making, but the bomb's incessant ticking has its own effect on the stability of securities markets.⁸ A consideration of the actions of the Federal Reserve in recent years aimed at dealing with so-called mini-crashes in the financial sector provides a further basis for assessing the prospects of centrally produced macroeconomic stability. From the narrow perspective of the financial sector the issues of malady and remedy look deceptively like those identified by Keynes: market maladies and extramarket remedies. An activist central bank is seemingly justified by its indispensable role in taming an otherwise wild financial sector. But a fuller understanding of the situation suggests that it is an unbridled Treasury rather than unbridled capitalism that lies at the root of the economy's current problems. And it is the Federal Reserve—its very existence—that removed the bridle. On this understanding, the malady and remedy are both in the extramarket category, but the diagnosis and prescription are not as simple as the Monetarists would have us believe.

Increasingly, the significance of the Federal Reserve in the context of the macroeconomy derives from its ability to monetize government debt. This is not to say that the actual rate of debt monetization dominates the Federal Reserve's current agenda but rather that the very potential for debt monetization is taking on increasing significance. How has the federal government been able to get away with such a chronically and conspicuously large budgetary imbalance—and with no sign of meaningful fiscal reform—without subjecting itself to the substantial penalty imposed automatically by credit markets? Why is there no default-risk premium on Treasury bills? Excessive debt accumulated by individuals, corporations, or even municipalities is eventually dealt with when the borrowers lose their creditworthiness and face prohibitive rates of interest. This salutary aspect of the market process is short-circuited in the case of Treasury debt by the very existence of a central

⁸There are a number of books written in the spirit of *Bankruptcy 1995* (1992) offering calculations of one sort or another about when the debt bomb will blow. Will it be when interest payments dominate the growth path of the debt? Or when interest payments exceed tax revenues? Calculations based on these and related eventualities are almost surely irrelevant. In informal discussion, I have designated all such calculations as establishing what I define to be the "Gore Point"—the point at which even Al Gore perceives the debt as a problem. (A colleague has suggested an equally apt name the "Barro Point," in honor of Robert Barro, who persistently downplays all the worries about government indebtedness.) The important point here is that financial markets do not await the education of Al Gore. Much of the instability currently observed on Wall Street is attributable to the chronically large debt and deficit.

bank. The Federal Reserve in its standby capacity as a buyer of government debt keeps the default-risk premium off Treasury bills. The potential for debt monetization allows federal indebtedness to rise unchecked to levels that would have been thought fanciful only a few administrations back and to remain high and rising into the foreseeable future.

The *potential* for debt monetization, critical for maintaining an uneasy balance between economic and political reality, gives rise to speculation about the timing and extent of *actual* debt monetization. At issue here are prospective movements, possibly dramatic ones, in the inflation rate, interest rates, and exchange rates, which in turn can have dramatic effects in securities markets. The attractiveness of securities can be differentially affected by the inflation that would result from actual debt monetization or by the movements in exchange rates that reflect the Treasury's greater or lesser reliance on foreign credit markets or by movements in interest rates brought about by changes in the Treasury's domestic borrowing. At some point, uncertainties about the timing and extent of debt monetization may dominate securities markets. In this case, the dense fog that drives our travelers off the economic highway and into the rest areas is not inherent in the market economy at all but rather is emitted by the Fed-backed Treasury.

It has become conventional wisdom in recent years that there is some link (though a poorly defined one) between chronically high budgetary deficits and instability of securities markets (Feldstein 1991, p. 8 and *passim*).⁹ And it is taken for granted that it is the Federal Reserve's responsibility to deal with that instability, providing on a timely basis whatever liquidity is demanded so as to keep the occasional sharp declines of security prices, the mini-crashes, from affecting the performance of the macroeconomy. The implicit objective, here, seems to be that of building a firewall between the financial sector and the real economy, allowing both to lead their separate lives. Ironically, it is largely the existence of the Federal Reserve—its potential for debt monetization—that enables the Treasury to borrow almost limitlessly, thus creating the very instability that is to be kept in check by that same Federal Reserve.

Short-term success of the Federal Reserve in maintaining the firewall between the financial and real economy depends critically on the wisdom and credibility of the Federal Reserve Chairman. Prospects for

⁹This is not to suggest that deficit-induced instabilities are the only macroeconomically significant ones. Instabilities emanating directly from the Federal Reserve and instabilities associated with perverse banking regulations and deposit-insurance pricing also have a claim on our attention. But, arguably, the deficit-induced instabilities deserve more attention than they have so far received. See Garrison (1993 and 1994).

longer-term success are problematic despite—or possibly because of—a sequence of short-term successes. Considerations of the nature of the Federal Reserve's role in the context of possibly volatile swings in the demand for liquidity suggest that continued central management of the economy's money supply does not offer the best hope for macroeconomic stability.

Suppose that the Treasury or the White House urges that the Federal Reserve become more accommodating and that the Federal Reserve Chairman expresses reluctance. Will the urgings get more intense? Will the reluctance fade? Speculation about the ultimate outcome will likely show up on Wall Street as an increased trading volume and an increased volatility of security prices. Traders who have little confidence in their own guesses about a possible change in the Federal Reserve's policy stance are likely to get out of the market. Securities prices weaken as these traders begin to liquidate, causing others to follow suit. Now, even those traders who do have guesses about the Federal Reserve begin guessing instead about the market's reaction to the uncertainty. The scramble to get out of the market manifests itself as a liquidity crisis. Abstracting from the fact that this instability has its origins in extramarket forces, we notice that the nature of this destabilizing speculation is exactly as described by Keynes (1936, pp. 153–58).

In dealing with the liquidity crisis, the Federal Reserve is immediately pitted against itself. It must expand the money supply to accommodate the increased demands for liquidity—and by the right amount in a timely fashion—while maintaining its credibility that it will not expand the money supply in response to the urgings from the White House. Fedwatchers are going to need some tea leaves here to determine just exactly what the Federal Reserve is and is not doing. Once again, the equation of exchange provides a sound basis for sorting it all out. M is being increased to offset a downward movement in V . If the increase in M is too little, the net downward movement in MV will result in the dreaded deflationary pressures which will impinge only partly on P and hence partly on Q . The Federal Reserve's firewall is too weak; the liquidity crisis spills over into the real economy. If the increase in M is too great, then, willy-nilly, the Federal Reserve is succumbing to the urgings of the executive branch to further accommodate the Treasury's borrowing. The extent of the accommodation, as measured by the net upward movement in MV , will in time show up as inflation, which was one of the prospective eventualities that underlay the speculation and the liquidity crisis.

As complicated and convoluted as this reckoning is, it constitutes only half of the story. Removal of the liquidity from the financial market in a timely manner is as important as its timely injection. The failure

of the Federal Reserve to move against an increasing V that characterizes the end of the liquidity crisis accommodates the Treasury and puts upward pressure on prices. Possibly more critical are the repercussions of the excess liquidity in international money markets. Overaccommodation can weaken the dollar. If this weakness is perceived as the beginning of a trend, the result may be heavy selling of dollars and dollar-denominated assets. Thus, a botched attempt to deal with a liquidity crisis can provoke a currency crisis. The Federal Reserve must somehow defend the real economy against this double-edged sword.¹⁰

The Federal Reserve may be allowed some scope for error. The same difficulties that it faces in knowing just what to do and just when to do it provide a shroud of uncertainty, even after the fact, about just what it did—and all the more so about what it intended to do. But several considerations combine to suggest that, in the long run, the Federal Reserve is playing against high odds.

First, right or wrong, the financial markets will make their moves ahead of the Federal Reserve. Changes in the demand for liquidity and in the strength of the dollar are determined as much if not more by anticipations about what the Federal Reserve will do rather than what it has just done. This consideration is what gives great importance to the Chairman's credibility. And his credibility reflects more than his personal integrity and his reputation for reasonableness and consistency. It is affected as well by the economic constraints he faces and political pressures he feels.

Second, each episode will have characteristics of its own depending upon all the contemporaneous political and economic factors. Goals of the Federal Reserve over and above the particular goal of accommodating the Treasury serve as a background against which expectations are formed. The Federal Reserve may be pursuing a strategy of gradual monetary ease to promote more rapid economic growth and then subsequently a strategy of gradual monetary tightening to stave off inflationary pressures. It may be possible to maintain credibility while increasing the monetary aggregates at an accelerated rate in the first episode but not possible while reversing the direction of change (relative to trend-line monetary growth) in the second episode.

Third, even if the Federal Reserve generally wins its battles against liquidity crises, it will find that winning streaks are difficult to maintain indefinitely. And perversely, a sequence of wins can create

¹⁰The idea that the Federal Reserve's attempt to deal with a domestic liquidity crisis may trigger an international currency crisis in this way is drawn from Lawrence Summers' discussion of the "Macroeconomic Consequences of Financial Crises" in Feldstein, 1991, pp. 153–56.

a false sense of confidence on Wall Street that the Federal Reserve is always willing and able to deal effectively with liquidity crises. Such confidence might cause investors to maintain a generally lower level of liquidity in their portfolios than if they had serious doubts about the streak continuing. Lower liquidity levels generally can mean more dramatic increases in the demand for liquidity during a crisis. For the Federal Reserve, the winning streak gets increasingly more difficult to maintain.

Temporarily and partially offsetting all these reasons for pessimism about prospects for enduring macroeconomic stability is the widespread belief that the particular individuals that have served as Federal Reserve Chairman are “geniuses.” Dating from the summer of 1979 Paul Volcker and, after him, Alan Greenspan have risen to the occasion whenever crisis threatened. It may indeed be difficult to name two other individuals who could have done better. “Genius” might involve overstatement; “seasoned,” “savvy,” and “nimble,” may be more to the point. But there is a greater point to be made here. Any governmental institution whose success depends critically on the caliber of the individual in charge cannot be considered a lasting source of stability for the economy. Even geniuses can err. More importantly, in some episodes where expectations turn pessimistic, the monetary ease needed to deal with a liquidity crisis may be more than enough to trigger a currency crisis. Foreign and domestic traders may leave no room for the Federal Reserve Chairman to exercise his genius. And further, geniuses are not necessarily succeeded by geniuses. Volcker served two four-year terms; Greenspan has begun his third term after an unsuspenseful reappointment in early 1996—which had the effect of postponing speculation for another four years. How much confidence will Wall Street have in Greenspan’s turn-of-the-century successor? How much confidence will it have in the Federal Reserve in the days or weeks before a successor is named? Suppose that the Treasury is putting pressure on the Federal Reserve for greater accommodation—possibly because our trading partners are reluctant to extend our government further credit until they know who is replacing Greenspan. What would happen to the demand for liquidity? And how would the lame-duck Federal Reserve Chairman respond so as to maintain his own credibility as well as that of his successor-to-be-named-later? Even mildly cynical or pessimistic answers to these questions may suggest that this financial crisis may burn through the firewall. The real economy would then become an innocent victim as the central bank attempts its extramarket remedy to the extramarket malady in the form of a fiscally irresponsible Treasury.

Free Banking as Both Prevention and Cure

The merits of free banking during periods of economic tranquility are identified on the basis of the theory of competition as applied to the banking industry and the experience provided by a key episode in nineteenth-century Scotland and more recent episodes involving other countries with partially free banking. Assessing the likely performance of free banking during twentieth-century financial crises in the United States necessarily involves some speculative reasoning. It is worth noting, however, that the most prominent nineteenth-century defender of free banking argued his case partly on the basis of the ability of competitive forces to “meet an incipient panic freely and generously” (Bagehot 1873, p. 104).

Whatever the problems and limitations inherent in free banking or in market economies generally, competition that characterizes a decentralized system wins out over the policy edicts of a central bank largely because of the absence of key perversities that are inherent in central control. The advantages of decentralization are partly in the form of prevention, partly in the form of cure.

One of the major sources of today’s macroeconomic instability, the excessive federal debt and deficits, would be largely absent under free banking. Without a central bank to keep the default-risk premium off Treasury bills, the federal government, like overextended firms and even fiscally irresponsible municipalities, would have had to deal with its fiscal imbalance long ago. Free banking, which is free not to monetize Treasury debt, could accomplish what debt-limitation ceilings, the Gramm–Rudman deficit-reduction plan, or even a balanced-budget amendment cannot accomplish. Without a chronically high and growing debt and the attendant speculation about the changing particulars of deficit accommodation, financial crises are less likely to occur.

If a financial crisis does occur, the provision of supernormal amounts of liquidity is forthcoming under free banking—but without the destabilizing speculation about the particular movements in the money supply. Questions about the “will” or “intent”—or “genius”—of the banking system as a whole simply do not arise. The supply of liquidity automatically follows demand upward during the financial crisis and downward as crisis conditions fade. It is true that some banks will be more responsive than others at meeting the occasional supernormal demands for liquidity. One of the beneficial aspects of competition in any sector of the economy is that those firms who best satisfy ever-changing demands prosper relative to their competition and are thus put in charge of greater resources. With free banking, then, success breeds success. A sequence of crises gives increased responsibility to those very banks that are best at dealing with crises.

To this point the advantages of free banking over central banking are set out in terms of the likelihood of our needing a firewall between the financial and real sectors of the economy and the ability of each banking institution actually to provide that firewall. The firewall metaphor, however, presumes that no systematic adjustments are needed in the real economy. But it is entirely possible and even likely that whatever caused the crisis conditions to prevail in the financial sector also caused non-financial resources to be misallocated. Simultaneous financial and real crises, as might be brought about by the ill-conceived policies of an administration bent on growing the economy, could not be quelled by a firewall. Quite to the contrary, the reallocation of resources in the economy would require a well-functioning market process, which includes movements in resources that reflect movements in securities prices. Here, the implicit monetary rule observed by free banking takes on a special significance. Movements on the lefthand side of the equation of exchange (an increasing V) are effectively countered; movements on the righthand side (in the p s and hence in P) are not. If the economy's real sector is out of balance, it needs help from the financial sector to regain its balance. In such circumstances, "firewall" is the wrong metaphor; "penny in the fusebox" would be more accurate. Only free banking can allow the financial sector to guide the real sector while preventing the demands for liquidity from degrading the market's performance.

A Summary View

In the Keynesian view, the central bank is a part of an extramarket remedy to a market malady. Investment markets are inherently unstable; government control of the economy's money supply is an important element in macroeconomic stabilization policy. The case against central banking—and for free banking—reverses the characterization of both remedy and malady. Free banking is a part of a market remedy to an extramarket malady. Even this stark reversal understates the case for free banking. It would remain valid even if we take the dramatic and chronic fiscal irresponsibility of the Treasury as given. Periodic crises that will inevitably occur in such a debt-ridden economic environment would be more ably countered by the market forces of free banking than by the policy moves of a central bank. But the extent of the Treasury's fiscal irresponsibility is itself dependent upon whether the Treasury can count on an accommodating central bank. Free banking limits the scope of this potential source of instability while at the same time enhancing the market's ability to deal with whatever instabilities that may persist.

References

- Bagehot, Walter. 1873. *Lombard Street: A Description of the Money Market*. London: Henry S. King.
- Ball, L., G. Mankiw, and D. Romer. 1988. "The New Keynesian Economics and the Output-Inflation Trade-off." *Brookings Papers on Economic Activity* 1: 1–65.
- Dowd, Kevin, ed. 1992. *The Experience with Free Banking*. London: Routledge.
- Feldstein, Martin, ed. 1991. *The Risk of Economic Crisis*. Chicago: University of Chicago Press.
- Figgie, Harry E., Jr. 1992. *Bankruptcy 1995: The Coming Collapse of America and How to Stop It*. Boston: Little, Brown.
- Friedman, M. 1969. *The Optimum Quantity of Money and Other Essays*. Chicago: Aldine.
- . 1970. *The Counter-Revolution in Monetary Theory*. London: Institute of Economic Affairs.
- Garrison, Roger W. 1993. "Public-Sector Deficits and Private-Sector Performance." In Lawrence H. White, ed. *The Crisis in American Banking*. New York: New York University Press. Pp. 29–54.
- . 1994. "The Federal Reserve: Then and Now." *Review of Austrian Economics* 8 (1): 3–19.
- Keynes, J. 1936. *The General Theory of Employment, Interest, and Money*. New York: Harcourt, Brace.
- Krugman, Paul. 1994. *Peddling Prosperity: Economic Sense and Nonsense in the Age of Diminished Expectations*. New York: W. W. Norton.
- Rothbard, Murray N. 1994. *The Case Against the Fed*. Auburn, Ala.: Ludwig von Mises Institute.
- Selgin, George A. 1988. *The Theory of Free Banking: Money Supply under Competitive Note Issue*. Totowa, N.J.: Roman and Littlefield.
- . 1991. "Monetary Equilibrium and the 'Productivity Norm' of Price-Level Policy." In Richard M. Ebeling, ed., *Austrian Economics: Perspectives on the Past and Prospects for the Future*. Hillsdale, Mich.: Hillsdale College Press. Pp. 433–64.
- Warburton, C. 1966. *Depression, Inflation, and Monetary Policies: Selected Papers: 1945–53*. Baltimore: Johns Hopkins University Press.
- White, Lawrence H. 1992. "Mises on Free Banking and Fractional Reserves." In John W. Robbins and Mark Spangler, eds. *A Man of Principle: Essays in Honor of Hans F. Sennholz*. Grove City, Penn.: Grove City College Press. Pp. 517–33.
- . 1984. *Free Banking in Britain: Theory, Experience, and Debate, 1800–1845*. New York: Cambridge University Press.
- Yeager, Leland B. 1986. "The Significance of Monetary Disequilibrium." *Cato Journal* 6 (2): 369–99.

Who Owes What, and To Whom? Public Debt, Ricardian Equivalence, and Governmental Form

Richard E. Wagner

The postwar literature on public debt has been concerned primarily with two related issues: (1) whether public debt allows the cost of government to be shifted forward onto future generations and (2) whether the creation of public debt involves a positive net wealth effect. The development of this literature can be portrayed chronologically by three works, even though many authors have contributed to that literature.¹ Lerner (1948) advanced the thesis that the burden of the debt rests upon people at the time the debt is created, as illustrated by the aphorism “we owe it to ourselves.” In sharp contrast, Buchanan (1958) argued that public debt allowed people in the present to shift the cost of government onto people in the future.² Barro (1974) denied the ability of public debt to transfer cost forward in time because, with intergenerational altruism, an increase in debt would be accompanied by an increase in saving to pay the future taxes required to service the debt. At the same time, Barro also denied the effectiveness of fiscal policy by that very fact: debt-financed government services would not have the stimulatory impact portrayed in the postwar Keynesian models, because the increased government spending would be offset by increased private saving necessary to provide the means to service the debt.

In this paper I accept the Ricardian argument that debt is just taxation by another name, both as an analytical point of departure and as a condition of satisfactory modeling. I also explain why the theory of public debt must be applied differently in democratic political settings than in the authoritarian setting that provides the effective backdrop

*Richard E. Wagner is Harris Professor of Economics at George Mason University.

¹For a survey of this literature, along with some effort to reconcile the various strands, see Vaughn and Wagner (1992).

²For a good sample of this controversy between these two positions, see the essays in Ferguson, ed. (1964).

for most of the postwar literature. To take a public debt of \$4 trillion for a population of 250 million, and to say that each person in the land has, on average, a debt of \$16,000 is an arithmetical truism that obscures rather than clarifies thought about public debt within a democratic state. If we ask *to whom* public debt is owed, we would not answer “ourselves,” but would answer “bondholders.” If we ask *who* it is that owes the debt, is “we” the right answer in any but a purely formal sense? Two people may agree to go for a ride, in which case they would no doubt each say that “we” went for a ride. But if one of them kidnapped the other, at least one of them would refuse to say “we” went for a ride. With regard to public debt, the question of “*who* owes what” cannot be addressed outside some framework for political or fiscal choice.

Personal Debt, Sovereign Debt, and Ricardian Equivalence

The use of deficit finance in place of current taxation represents the substitution of future taxation for present taxation, and the present value of those future taxes will equal the amount of the deficit being financed. This is simply a matter of arithmetic. An *aggregate* balance sheet can show no change in net worth, because all that has happened is that a short-term liability has been transformed into a long-term liability of the same present value. The aphorism “we owe it to ourselves,” is simply an identity within any aggregate accounting system. In any aggregation over balance sheets, assets must equal liabilities. The value of home mortgage debt held by creditors must equal the liabilities of mortgagees. Viewed in the aggregate, it would be correct to say “we owe mortgage debt to ourselves.” But it would not be correct to infer from this statement that payments on mortgage debt are simply transfer payments and not payments for services rendered.

Ricardian equivalence must provide a point of departure for any analysis of public debt, as well as serving as a necessary constraint on any effort at *aggregate* modeling. But it does not follow from this simple arithmetic that the choice between debt and taxation is subject to some invariance proposition to the effect that nothing depends upon or is affected by the choice between debt and taxation. The choice between debt and taxation can matter for *particular* people, and, indeed, these differences are central for any effort to understand the creation of public debt in the first place. Aggregate equivalency must not be confused with a proposition of behavioral invariance to particular institutional conditions. It does not imply that choices between debt and taxation will be invariant to the institutional setting within which such choices are made.³ To the contrary, different institutional settings can lead to different

³The importance of institutional settings for debt analysis is explored in Wagner (1986).

fiscal and budgetary choices, despite the underlying constraint implied by Ricardian equivalence. I shall consider this proposition briefly here for personal and sovereign debt, before considering democratic debt in the remainder of this paper.

Ricardian equivalence clearly holds for personal debt. Someone who borrows to buy a car does not become wealthier than he would have been by paying cash. The reduction in cash that would have been required is equivalent to the present value of the liability for amortization payments when the car is financed by borrowing. Some people may prefer loan finance to cash finance, perhaps because loan finance allows the buyer to achieve a smoother time path of total consumption than would be possible with cash finance. Nonetheless, loan finance does not allow someone to escape his intertemporal budget constraint. It is not a source of wealth.

The Ricardian character of personal loan finance stems from the institutional setting within which credit markets operate. A borrower might well like to derive a positive wealth effect from borrowing, but the lender has a strong interest to prevent such a wealth transfer. Collateral provisions and risk premiums are means to guard against such wealth effects. So too, for that matter, is the borrower's own interest in avoiding a bad reputation, at least so long as he is interested in future credit transactions. While personal borrowing has no wealth effect *ex ante*, there can be instances where it does have such an effect *ex post*. The borrower may become insolvent, or may die insolvent, with the lender receiving less than full payment in either case. Even in these cases, however, Ricardian equivalence still holds in the aggregate because the borrower's gain is offset by the lender's loss.

This institutional setting for personal loan finance could be extended readily to state borrowing in a monarchical or other form of authoritarian regime. State debt in this setting is equivalent to personal loan finance when dealing with individuals. Indeed, in the case of a hereditary monarchy, full intergenerational altruism may be plausible, though it is not necessary. For even if such altruism were only partial, and even if the monarch's hold on his crown were tenuous, the Ricardian constraint would nonetheless hold as a central feature of credit markets. Debt finance could offer an anticipated increase in net worth to a sovereign only in the event he had superior information about his intention to migrate or to die than creditors possessed. In such circumstances, the sovereign might choose loan finance over tax or cash finance as a means of appropriating the assets of creditors. The sovereign's ability to do this, however, is limited by the interests of creditors to avoid this appropriation, and so is possible only to the extent informational asymmetries are present. And even should cases of

such appropriation result, the Ricardian constraint must again hold in the aggregate: the sovereign's increase in net worth must be offset by a decrease in the net worth of his creditors. Ricardian equivalence must always hold over some appropriately defined aggregate.

The theory of autocratic or sovereign loan finance does not really look much different from the theory of personal loan finance.⁴ Differences there are, but the similarities are dominant. An inability to bring suit against a monarch for a failure to pay debts is a difference whose significance is easy to exaggerate, just as it is easy to exaggerate the importance of police and courts in explaining why people generally adhere to personal credit contracts. A monarch faces the problem of attracting credit in the first place, and the willingness of lenders to lend varies directly with the strengths of their beliefs that the monarch will pay his debts. A sovereign's ability to appropriate the assets of creditors through default or repudiation is limited by the interests of creditors in avoiding this appropriation.

More than this, even in monarchies and other forms of absolutism, extra-legal means of contract enforcement are available. There is no such thing as an unconstrained sovereign (Tullock 1987). Among other things, unhappy creditors can serve as threats to a reigning despot through their ability to organize coups, to plot assassinations, and the like. An indebted king will have more options and resources than an indebted peasant, though he will surely also have larger debts which, in turn, will help to marshal stronger opposition to repudiation. Moreover, a peasant's debts might be so small as to lead a creditor to decide that it is not worthwhile to pursue a legal action for recovery of what is owed. In any case, the differences between personal and sovereign debt is surely more quantitative than categorical.

External Debt, Internal Debt, and State Borrowing

Personal or sovereign debt can only be external, for it makes no sense to speak of a person or sovereign as borrowing from himself. A considerable body of scholarship portrays democratic debt as if it were the debt of a person or sovereign. Barro (1979, 1989), for instance, explains public debt in terms of the utility-maximizing choice of a representative citizen where the government is faced with exogenous shocks to anticipated revenues or planned spending, with those shocks usually described as recessions or wars. If tax rates are varied in response to such shocks to maintain a balanced budget, the excess burden of taxation will be larger than it would be if taxation were held constant at that level which produced long-term budget balance. According to this tax-smoothing explanation, public debt

⁴On the economic theory of sovereign debt, see, for instance, Eaton and Gersovitz (1981) and Grossman and Van Huyck (1988).

smooths shock-induced variations in tax rates, and thereby minimizes the excess burden associated with taxation. This formulation represents a public sector counterpart to the permanent income hypothesis, with transitory deviations from normalcy going into public sector saving, either negative via debt creation or positive via debt reduction. While this formulation builds upon a good deal of intuitive plausibility grounded in personal finance, it is nonetheless quite problematical precisely because of its personification of the state.

For one thing, this standard reasoning in support of public borrowing to finance extraordinary expenditure does not explain why the result is a theory of public debt and not a theory of public credit. Even if the analogy with the permanent income hypothesis is maintained, it does not follow that the state will be a net debtor. It could just as well be a net creditor. Borrowing to finance unanticipated decreases in revenue or increases in expenditure is only one of two possible options. The other option is to finance those unanticipated deviations from some type of reserve fund, in which case the state would be a net creditor.

It may be readily acknowledged that extraordinary circumstances may periodically place unusually heavy demands on governmentally-provided services. Yet this acknowledgement does not support a model of public borrowing over one of government as a supplier of credit. The central point in either case is that, in present value terms, the actual tax rate exceeds the rate required to finance ordinary expenditure by an amount sufficient to cover the extraordinary expenditure.⁵ Whether a government would borrow or lend would depend primarily on the historical timing of recessions or wars. If one occurs at the beginning of a regime, it would have to be financed by borrowing. Otherwise, it could be financed through a fund created by previous surpluses. In any case, an aggregation over governments would show no net governmental debt. To be sure, in representative democracies there may well be reasons for an asymmetry between budget surpluses and budget deficits, with surpluses having less political value.⁶ Recognition that the claims of extraordinary finance cannot be used to explain why states are borrowers rather than creditors suggests, in turn, the merits of exploring a less aggregative conceptualization of public borrowing, recognizing all the while that the Ricardian constraint must operate in the aggregate.

Moreover, explanations of state borrowing through such analogies with permanent income notions as the tax smoothing hypothesis require that all public debt be held externally. Yet the preponderant share of public debt is held internally, and tax smoothing cannot explain internal

⁵Or to cover reductions in tax receipts below normal levels in the case of recessions.

⁶This theme is developed in Buchanan and Wagner (1977).

debt finance. With internal public debt, some people lend and others borrow. People who buy bonds make what would otherwise have been the tax payments of those who do not buy the bonds. The borrowers might be smoothing their tax payments, but the lenders are amplifying the variability of their tax payments. With internal public debt, the state acts as an intermediary between those people who advance credit and those people who defer their tax payments until some later period.

Democratic Debt and the Intermediary State

To say that the government owes \$4 trillion or must pay \$200 billion in interest on its debt is not to acknowledge some organizational obligation, for that would be to personify what cannot be personified. Rather, it is a reductionist manner of saying that some people owe \$4 trillion to other people, and that the annual interest component of that obligation is \$200 billion. The government is in essentially the same position of any other financial intermediary, with the intermediation taking place between the borrowers who thus are paying less in current taxes and the lenders who are making those tax payments instead. With ordinary financial intermediaries, it is straightforward to say who owes what to whom. It is not so straightforward, however, for public debt in a democracy. The aphorism “we owe it to ourselves” offers no guidance in this respect. It is even wrong, in that the “ourselves” can be replaced by the bondholders. But who is it, precisely, who does the owing?

One question that arises immediately is why the state would act as an intermediary. Why does public debt substitute for taxation? The predominant response to this question has involved the presumption that this substitution takes place when it is economically efficient along Paretian or Kaldor–Hicks lines. An additional tax burden is imposed in a setting where, say, 90 percent of the population would prefer to defer payment at prevailing interest rates, and where the remaining 10 percent is willing to lend at those rates to the other 90 percent. Why not assign the tax liabilities in the present, and let people work out whatever credit arrangements they choose through ordinary market processes? Private debt would replace public debt. The efficiency-based explanation for state intermediation through public debt is that government can borrow more cheaply than could the individuals who would otherwise resort to private lending sources.⁷ State intermediation offers potential gains from trade for all participants, when compared against market-based intermediation. On the one side, people who supply credit at the government's interest rate reveal their willingness to do so. On the other side, it would seem to take no genius to determine that the remainder

⁷See, for instance, de Viti de Marco (1936, pp. 377–98).

of the citizenry, who otherwise have to borrow at commercial rates, would prefer to borrow at the government's lower rate.

If this is all there were to the matter, government would be a superior intermediary to market-organized firms. In this case, the lower rate of interest on public debt would represent a genuine cost advantage over transactions organized through market-based intermediaries. Socialized intermediation would be superior to privatized intermediation. The lower rate of interest on public debt is undeniable, as a simple inspection of the financial pages of any newspaper will show. It does not follow, however, that there are any efficiency gains from the state supply of financial intermediation. There is a neglected cost to state-organized intermediation, which is a cost that is borne by taxpayers because the state uses its power to impose unlimited liability on taxpayers.⁸

It is accurate to say that government bonds carry a lower rate of interest than corporate bonds because bondholders believe that the government bonds are safer. This safety, however, resides not in greater governmental efficiency, but in the unlimited liability that taxpayers face, as contrasted to the limited liability of corporate shareholders. If corporate bonds were sold under unlimited liability, where officers and directors were personally liable for claims, they would offer lower interest rates than they now carry, perhaps even lower rates than offered by state debt. Alternatively, if government operated by principles of limited liability, bondholders would bear risk that is now borne by taxpayers, and the government's borrowing rate would rise, perhaps above that which is privately available. The lower interest rate on public debt may represent not some genuine opportunity but rather be a feature of the unlimited liability character of governmental claims. As a result, the resort to public debt may result not because everyone gains, but because some gain to the detriment of others, with public debt serving as a vehicle of wealth redistribution.

Public Debt in a Concordant Democracy

While it seems relatively straightforward to conclude that public debt is owed to the state's creditors, only in a purely formal sense can it be concluded that the general citizenry constitute as a collective body the state's debtors. The question of who truly owes public debt depends upon the character of a political system and its institutions. An important question in this respect is whether that borrowing is done voluntarily or forcefully. Is it that the net debtors agreed to the transaction? Or is it that they were compelled, and hence became *forced debtors*? The typical presumption in the literature on public debt is that the transactions

⁸This point is articulated in Mises (1966, pp. 225–28).

were voluntary. To be sure, this presumption is often left only implicit, as in the sovereign debt literature where only a single person is involved, or in the related notions involving some representative citizen. Whether the intermediation between debtors and creditors that public debt entails involves free debtors or forced debtors is to an important extent a matter of whether the constitutional framework within which fiscal choices are made can be characterized as *concordant* or *factional*.

It is certainly possible to imagine fiscal choices being made in an essentially concordant constitutional system, as illustrated by various models of the optimal supply of public goods. Consider a Wicksellian fiscal system, where fiscal choices would be made under unanimity or something close to it.⁹ Borrowing would be agreeable to all, net debtors and net creditors alike. The only difference in this case from the usual presentations of Wicksellian-based models is that a secondary market for personal tax obligations would arise. Tax liabilities would be assigned in the present, under unanimity or close to it, only the state would act as an intermediary to bring together those who would prefer to defer their tax payments and those who are willing to lend.

Public debt clearly would fit within the Wicksellian framework if the future tax liabilities that debt finance entailed were assigned in the present. This would be Ricardian equivalence reduced to the level of the individual participants in the current choice to replace current taxation with future taxation, for those future tax obligations would be assigned to particular individuals in the present. Table 1 illustrates this point in a simple fashion, and in a manner that will be used for a different purpose in the next section. Assume there are three voters and two periods. In the absence of government debt, total taxes and government spending are \$300 in each period, with each voter paying \$100 in each period. Alternatively, suppose taxes are reduced by 10 percent, with the remainder supplied by a \$30 issue of public debt where V_1 buys the government bonds. Assuming a zero rate of interest to simplify the illustration, total tax collections rise to \$330 in the next period, \$300 of which go to finance state services while the remaining \$30 is paid to V_1 to service the debt. The present value of aggregate tax collections over the two periods, assuming a zero rate of discount, is \$600 with or without debt. Similarly, the present value of each person's tax liability is \$200, with and without debt. With debt V_2 and V_3 each borrow \$10 from V_1 , and repay their debts in the next period.

While debt finance can in principle fit within the Wicksellian, concordant framework, it is questionable whether it can do so in practice.

⁹Wicksell is often summarized in this regard as supporting a rule of near-unanimity in collective choice. This is true, but there also is a supporting institutional framework that complements the principle of unanimity, and which is discussed in Wagner (1988).

Table 1
Illustration of Public Debt Options

Without Debt	Period 1	Period 2
V ₁	\$100	\$100
V ₂	100	100
V ₃	100	100
Total	300	300
With Debt		
V ₁	\$90 + 30	\$110 - 30
V ₂	90	110
V ₃	90	110
Total	270 + 30	330 - 30

In the absence of such a present assignment of future tax obligations, the actual distribution of those future obligations that result from the present choice to borrow will be contingent upon a wide variety of circumstances. Among other things, these circumstances include the future economic standing and position of people in a future period when those taxes become due, and also include all of the possible political adventures that may change the tax code over the intervening years. Circumstances could be imagined in which the contingent feature of the tax liability that debt issue represented was expectationally neutral, in that the present value of anticipated tax liability was the same for everyone with and without debt. Indeed, within a Wicksellian framework for fiscal choice, a substitution of the future taxation that public debt represents for current taxation would be approved only if people had neutral expectations as to the impact of the contingent character of the future liability that public debt creates. Otherwise, there would be people who expected public debt to increase their tax obligations relative to current tax finance, and so in turn would oppose the proposed debt finance, even if they might support the proposal under current tax finance.

Public Debt in a Factionated Democracy

What about public debt within a factional constitutional framework? By factional I mean a constitutional system that fails to control what

Madison called in *Federalist* No. 10 the violence of faction, and which conforms to rent seeking and churning in the contemporary literature on public choice and constitutional economics. There are many particular ways to model a factional constitutional system. In all such cases, some people are able to enrich themselves by securing increased spending on desired programs, paid for by taxes imposed on other people.

How might the elimination of the constitutional constraint on public borrowing affect the resulting budgetary choice? For the option of deficit finance to have an impact on budgetary choices, it is necessary that the introduction of that option expand the opportunities available to some decisive subset of the population. Deficit finance would have to lower the cost of budgetary choices to decisive individuals and coalitions, as compared with tax finance. By virtue of the Ricardian theorem, the aggregate present value of future taxes must equal the amount of the budget deficit. But it does not follow that such present-value equivalence holds across individuals, which means in turn that the consequences for budgetary choice will depend on the way in which fiscal institutions shape and constrain processes of budgetary choice.

The situation portrayed in Table 1 can be used to illustrate some central features of factional public debt. There are several ways that the creation of public debt can change the cost of government to voters in period 1, while maintaining the Ricardian equivalence built into Table 1 in the aggregate. One way is through recognition that the identities of the people portrayed in Table 1 change with the passing of time. For instance, V_3 in period 1 may be elderly and without heirs, with his place in period 2 taken by a new entrant, who simply faces a tax burden of \$110, \$10 of which goes for interest on the public debt. Alternatively, V_3 might be middle-aged in period 1 and be in retirement in period 2 and out of the labor force, with his place taken by a new entrant in the labor force. For people in such positions as these, debt finance is less costly than tax finance.

Within a factional system, public debt is one of the instruments of wealth redistribution and government expansion. Consequently, public debt would not represent some agreement between net debtors and net creditors. Net creditors would clearly agree to hold the debt, at an interest rate that reflects their assessment of risk. Not all net debtors, however, would have agreed to the transaction. Some would have, namely those who gained through budgetary expansion, when they would not have supported such expansion under tax finance. Public spending and debt would expand beyond what a significant share of the population would have agreed to under some Wicksellian-type constitution. Creditors gain from the operation, as does that part of the net debtor population that nonetheless are net gainers from the debt-financed expansion in government. There will also be forced debtors who would have preferred tax finance, along with the

smaller public sector that would have resulted. For instance, someone with a number of children and a strong bequest motive will lose through public debt, as personal consumption is reduced to provide the fund to pay the resulting added taxes imposed on his heirs.

There are many dimensions along which this separation among net debtors might occur. One involves differences in the degree of intergenerational altruism among people.¹⁰ In the presence of the constitutional prohibition on deficit finance, variations in intergenerational altruism will have no impact on budgetary choice. But when borrowing is possible, the no-debt outcome will be disturbed. For borrowing reduces the relative cost of government services to people the weaker their intergenerational altruism, and provides a vehicle for leaving negative bequests. In a simple median voter model, the budgetary choice will be controlled by the person whose intergenerational altruism is median within the population. The introduction of a deficit financing option will lead to an expansion in the size of government because it reduces the cost of government to the median voter. A new budgetary equilibrium will be established where, for the median voter, the marginal value of added public output equals his marginal cost through deficit finance.

Other models could give descriptively different but analytically similar results. For instance, a ruling political party could be characterized as seeking to expand different tax sources so as to equalize political resistance at the respective revenue margins. The introduction of a debt option lowers marginal political cost. This leads to deficit finance, and the more fully debt is used the higher becomes the political cost of deficit finance. The political pressures from different revenue sources will be equalized at the relevant political margins, where the future taxes represented by debt finance encounter the same political resistance encountered by present taxes. The Ricardian proposition must hold as a condition of political equilibrium, for otherwise there will be a shift in the mix of tax instruments toward those that entail lower political cost.

In any case, public debt is to a significant extent concerned with transferring wealth among people within the present, though it also exerts effects across time. Among other things, people with relatively weak bequest motives promote the use of public debt over taxation as a method of increasing their net wealth, while people with relatively strong bequest motives suffer a wealth loss through the larger-than-desired public sector that results. The creation of public debt does not increase aggregate wealth, but it does increase wealth for some people who are influential at the margins of budgetary choice, while reducing wealth for those who are on the losing side.

¹⁰See, for instance, Buchanan and Roback (1987) and Cukierman and Meltzer (1989).

Repudiation as Constitutional Restoration?

In a world of balanced budget finance, some people can be forced carriers as a natural product of a factional constitutional system. When deficit finance is allowed, people can likewise become forced debtors. These are people who would not have supported expansion in debt-financed programs, but who will be compelled to pay taxes to amortize that debt in future years. Government is expanded beyond concordant limits, as the result of a coalition between net creditors and those net debtors who gain from budgetary expansion. What remain are those net debtors who lose from budgetary expansion, and who find themselves saddled with future taxes to amortize the debt-financed spending that they would not have consented to in the first place.

A number of commentators have expressed concern that the debt-income ratio could rise to a point where people begin to lose confidence in the government's ability to service its debt.¹¹ This positive claim can be granted without coming to the conclusion that public debt should be frozen, as by the imposition of a balanced-budget requirement, or retired. Public debt can also be repudiated, either partially or wholly. In this regard, Rothbard (1962, pp. 881–83) argues not just against the view that “we owe it to ourselves,” but also argues in support of repudiation, as against either retirement of debt or freezing it at present levels. It is plain to see why state creditors and the winning borrowers within a system of factionated democracy would oppose any effort at debt repudiation. Not only would such repudiation erode gains that have been set in place by past fiscal operations, but also repudiation would curb sharply, if not eliminate entirely, the willingness of people to serve as state creditors. Repudiation would seem to increase the cost of participating in the factionated politics of shifting tax burdens onto losing coalitions through deficit finance, by increasing the costs to potential creditors from participating in the process. By reducing the willingness of people to become state creditors, repudiation would increase the cost of deficit finance and, hence, lower the size of government and the present value of future taxation.

To be sure, public debt is woven throughout our society, with many people being state creditors through their share in the holdings of pension funds, as well as in other ways. No doubt, many such people might be state creditors while still being net state debtors, even forced debtors at that. Repudiation would affect balance sheets throughout the land, and would probably leave few untouched. Taxation and monetization offer alternatives to repudiation that could achieve more-or-less the same end-state, though with considerable distributional differences among people.

¹¹See, for instance, the discussion in Spaventa (1988). In a related vein, see Steadman (1993) for a program of debt retirement.

Indeed, repudiation is a form of taxation on the holders of government bonds. It seems clear enough why politicians and all of those who think they are net gainers from the continual churning of the state machinery would support either taxation or monetization over repudiation.¹² Indeed, taxation and monetization have been used repeatedly in covering over some of the insolvencies that arise through the state's intermediation between creditors and debtors, under which the promises to pay to creditors exceed the promises not to tax debtors.¹³ Repudiation, however, would seem to strike at the very political process that countenances the creation of insolvent claims within the political marketplace. Should shifting political coalitions present an opportunity for repudiation, some quasi-constitutional assurance against a future status of being a forced borrower might be created through the long-term impact that repudiation would have upon potential state creditors.

Concluding Remarks

The economic analysis of public debt differs depending on the presumed political setting. It may be reasonable to characterize public debt choices within an authoritarian regime as being made by a single mind, but such a characterization is surely inapt for democratic regimes. A presumption of a representative citizen is simply incapable of characterizing public debt choices in complex societies where preferences and interests differ among people. To be sure, even in such societies Ricardian equivalence must hold in the aggregate, but this aggregate equivalence is irrelevant for human conduct in fiscal choice. Deficit finance injects a systematic differential among current citizens in the cost of public finance.

In this paper I have considered public debt as an alternative to tax finance. But within existing monetary institutions, public debt creation often serves as a disguised form of money creation. The possibility of inflationary finance opens up, in turn, new avenues along which deficit finance may serve as a means by which politically dominant groups are able to impose costs on others. A complete analysis of public debt within an interest-group approach to fiscal processes will clearly have to incorporate and integrate such monetary considerations, at least under prevailing central-banking institutions. Wherever such an analysis might lead, Ricardian equivalence will have to hold in the aggregate; yet such aggregative equivalence will be only a side show in the fiscal drama that public debt represents.

¹²The churning state is articulated nicely in de Jasay (1985).

¹³A tax rate of 10 percent in place at a time when debt is created involves as much a promise to let taxpayers keep 90 percent of their income as it involves a promise to pay creditors principal plus interest. An increase in the tax rate is as much a partial repudiation as is an erosion through inflation in the real value of payments to creditors.

References

- Barro, Robert J. 1974. "Are Government Bonds Net Wealth?" *Journal of Political Economy* 82: 1095–118.
- . 1979. "On the Determination of the Public Debt." *Journal of Political Economy* 87: 940–71.
- . 1989. "The Neoclassical Approach to Fiscal Policy." In Robert J. Barro, ed. *Modern Business Cycle Theory*. Cambridge: Harvard University Press. Pp. 178–235.
- Buchanan, James M. 1958. *Public Principles of Public Debt*. Homewood, Ill.: Richard D. Irwin.
- Buchanan, James M., and Jennifer Roback. 1987. "The Incidence and Effects of Public Debt in the Absence of Fiscal Illusion." *Public Finance Quarterly* 15 (January): 5–25.
- Buchanan, James M., and Richard W. Wagner. 1977. *Democracy in Deficit*. New York: Academic Press.
- Cukierman, Alex, and Allan H. Meltzer. 1989. "A Political Theory of Government Debt and Deficits in a Neo-Ricardian Framework." *American Economic Review* 79: 713–32.
- de Jasay, Anthony. 1985. *The State*. Oxford: Basil Blackwell.
- de Viti de Marco, Antonio. 1936. *First Principles of Public Finance*. New York: Harcourt Brace.
- Eaton, Jonathan, and Mark Gersovitz. 1981. "Debt with Potential Repudiation: Theoretical and Empirical Analysis." *Review of Economic Studies* 48: 289–309.
- Ferguson, James M., ed. 1964. *Public Debt and Future Generations*. Chapel Hill: University of North Carolina Press, 1964.
- Grossman, Herschel I., and John B. Van Huyck. 1988. "Sovereign Debt as a Contingent Claim: Excusable Default, Repudiation, and Reputation." *American Economic Review* 78: 1088–97.
- Lerner, Abba P. 1948. "The Burden of the National Debt." In Lloyd A. Metzler, ed. *Income, Employment, and Public Policy*. New York: W. W. Norton. Pp. 255–75.
- Mises, Ludwig von. 1966. *Human Action*. 3rd rev. ed. Chicago: Henry Regnery.
- Rothbard, Murray N. 1962. *Man, Economy, and State*. Los Angeles: Nash.
- Spaventa, Luigi. 1988. "Introduction: Is There a Public Debt Problem in Italy?" In Francesco Giavazzi and Luigi Spaventa, eds., *High Public Debt: The Italian Experience*. Cambridge: Cambridge University Press. Pp. 1–24.
- Steadman, Charles W. 1993. *The National Debt Conclusion*. Westport, Conn.: Praeger.
- Tullock, Gordon. 1987. *Autocracy*. Dordrecht: Martinus Nijhoff.
- Vaughn, Karen I., and Richard E. Wagner. 1992. "Public Debt Controversies: An Essay in Reconciliation." *Kyklos* 45 (1): 37–49.
- Wagner, Richard E. 1986. "Liability Rules, Fiscal Institutions, and Public Debt." In J. M. Buchanan, C. K. Rowley, and R. D. Tollison, eds. *Deficits*. Oxford: Basil Blackwell. Pp. 199–217.
- . 1988. "The Calculus of Consent: A Wicksellian Retrospective." *Public Choice* 56 (February): 153–66.

Reflections on the Misesian Legacy in Economics

Israel M. Kirzner

As I begin this paper for the issue of the *Review of Austrian Economics* published in honor of the memory of Murray N. Rothbard, my mind goes back over 40 years, to the first time that I met him. It was at the opening session of the Seminar in Economic Theory which Professor Mises conducted in the fall semester of 1954. That occasion was also my first meeting with Ludwig von Mises, and it is etched deeply in my memory. Two statements by Mises at that seminar meeting stand out in my recollection. One statement was his very opening substantive sentence that evening. "The market," Mises began, "is a process." (See also the statement in *Human Action* [1966, p. 257]: "The market is not a place, a thing, or a collective entity. The market is a process.")

Coming as I did from a rather spotty undergraduate training in economics (and mainly along Keynesian lines), Mises's statement, I recall, left me completely puzzled. I *had* thought of the market as a place, an arena for exchanges, as an abstract idea referring to voluntary exchange transactions. I could not fathom what on earth could be meant by the observation that the market is a *process*. I now, in retrospect, consider that all my subsequent training and research in economics, both before and after obtaining my doctorate under Mises, has consisted in learning to appreciate what it was that Mises meant by this assertion.

The second statement by Mises which stands out in my memory from that September 1954 evening, is a reference that Mises made to Murray Rothbard. Murray had, it appeared, recently completed a paper which Mises found to be excellent. He briefly but warmly complimented Murray on that piece of work, and expressed the hope and the prediction that Murray would continue to produce a great deal of future work of similar excellence. The years since 1954 have amply borne out

*Israel M. Kirzner is professor of economics at New York University.

The Review of Austrian Economics Vol. 9, No. 2 (1996): 143–54
ISSN 0889-3047

Mises's hope and his prediction. Murray Rothbard's output during these four decades has been prodigious. The breadth of his reading across so many disciplines has been breathtaking; his sheer energy in producing thousands of pages of published work has been stupendous. It is a privilege to contribute this paper to a memorial issue dedicated to the memory of Murray N. Rothbard.

My paper will have to do with the first of the two statements made by Mises at that 1954 seminar session. I will be taking issue with a certain tendency, present in a number of recent expositions of Mises's work, to de-emphasize (or even flatly to deny) the centrality of the idea of the market as a *process* in the Misesian system. I consider clarification concerning the character of the Misesian system to be of critical importance for the future direction of modern Austrian economics, and for its ability to contribute fruitfully to the restoration of economic understanding for the economics profession and for intelligent lay people at large. And this matter is also, of course, of fundamental importance in projecting an accurate overall view of Mises's own contributions. While I shall, in my argument, be taking issue with a number of relevant statements by Rothbard, I trust that the reader will appreciate that the purpose of this paper is simply to further that very Misesian legacy to which Rothbard dedicated his entire life's work as an economist. It is as a memorial to Murray Rothbard's consistency in this regard, and his willingness to bear formidable costs to his professional career in order not to compromise the honesty of his expositions, that this paper is being written. The purpose of any critical observations in this paper (whether directed at Rothbard or at others) is certainly not to stir up strife within the Austrian camp; quite the reverse. I am convinced that a clear, shared understanding of Mises's central vision can bring together all those who appreciate the intellectual content of the Misesian legacy. To contribute an attempt in this direction, in honor of the memory of an outstanding exponent of that legacy, is the purpose of this paper.

The Misesian Market Process

My own understanding of what Mises means when he describes the market as a process can be stated simply:

(1) Mises saw the market process as a continually *corrective* process driven and constituted by active entrepreneurial grasping of pure profits. "The essential fact is that it is the competition of profit-seeking entrepreneurs that does not tolerate the preservation of *false* prices of the factors of production. The activities of the entrepreneurs are the element that would bring about the unrealizable state of the evenly rotating economy if no further changes were to occur" (Mises 1966, pp. 337–38; emphasis in the original). The market process consists, that is,

in the continual correction of false prices that occurs in the course of entrepreneurial competition. If exogenous changes were not to occur, this corrective process would eventually lead to a price structure for factors of production and consumer goods, in which all entrepreneurial profit has been squeezed out. In the real world, at any given moment, factors of production are able to be purchased at false prices, prices which permit entrepreneurs to capture pure entrepreneurial profits. False prices are false in that they incorrectly reflect the relative urgency of consumer demand for the various alternative possible products that can be created with these factors. It is this discoordination between what *might* be produced and what in fact is being produced, which offers alert entrepreneurs opportunities for pure gain. "What makes profit emerge is the fact that the entrepreneur who judges the future prices of the products more correctly than other people do buys some or all of the factors of production at prices which, seen from the point of view of the future state of the market, are too low" (Mises [1951] 1962, p. 109). Entrepreneurs "are the first to understand that there is a discrepancy between what is done and what could be done." Their activity brings about a systematic adjustment of factor prices. They "bid higher prices for some factors of production and lower the prices of other factors of production by restricting their demand for them." Their activity also generates price adjustments for consumer goods. "In supplying the market with those consumers' goods in the sale of which the highest profits can be earned, they create a tendency toward a fall in their prices. In restricting the output of those consumers' goods the production of which does not offer chances for reaping profit, they bring about a tendency toward a rise in their prices. All these transformations go on ceaselessly and could stop only if the unrealizable conditions of the evenly rotating economy and of static equilibrium were to be attained" (Mises 1966, p. 336). All this ceaseless sequence of corrective price adjustments constitutes Mises's entrepreneurial market process.

(2) This Misesian corrective process from a false set of prices towards a set of fully mutually adjusted prices may be restated in the terms in which Hayek understood the market process to constitute a "discovery procedure" (Hayek [1968] 1978, chap. 12). "False" prices reflect the decisions of entrepreneurs who have not yet understood the correct implications of consumer preferences (present or future) for the relative values of resources today. The way in which entrepreneurial activity tends to correct such false prices is through their realization of the profit possibilities inherent in such false prices. Grasping these profit possibilities is the way in which entrepreneurs express their discoveries concerning the correct valuation of resources (and thus, in effect, concerning better ways in which resources can be

deployed in serving the preferences of consumers). The tendency which this entrepreneurial process generates towards equilibration is thus one of gradually enhanced mutual anticipation on the part of market participants. In the theoretical limit, in the hypothetical state of equilibrium in which no entrepreneurs would earn profit or suffer losses, we would be able to say that “all people . . . anticipate correctly the future state of the market” (Mises [1951] 1962, p. 108). Although it was Hayek, rather than Mises, who extensively articulated the nature of the market equilibrating process as one of gradually enhanced mutual knowledge, there can be no doubt that an interpretation of the Misesian process in terms of enhanced mutual knowledge is a valid one. Disequilibrium prices are “false” prices; as entrepreneurial profit taking nudges prices towards their correct levels, entrepreneurs have been led to more accurate anticipations concerning relevant future market configurations.

(3) What makes possible the entrepreneurially driven process of equilibration is active market *competition*. It is only the possibility of unrestricted entrepreneurial entry which permits more alert entrepreneurs to deploy their superior vision of the future in order to correct the misallocations of resources reflected in the false prices which characterize disequilibrium. It is the continual threat of such entry which tends to keep incumbent entrepreneurs alert and on their toes. The reason that Mises had little patience for the concept of perfect competition (see his approving reference to Hayek’s pioneering essay on this matter “The Meaning of Competition,” in Mises 1966, p. 278n), was that this concept can relate only to an already attained state of equilibrium. It has nothing to do with, and can throw no light upon, the competitive forces which drive the entrepreneurial market process. In deepening his (and our) understanding of the competitive process as consisting in a discovery procedure, Hayek was articulating insights that are, at the very least, thoroughly consistent with Mises’s own understanding of the dynamic entrepreneurial competition which, for Mises, constitutes the heart of this market process.

The Shared Understanding of Mises and Hayek on the Market Process

To draw attention, as we have in the preceding paragraphs, to the shared understanding that is apparent in Mises’s and Hayek’s treatment of the market process, is not to “homogenize” separate systems or “paradigms” of economic thought. Mises and Hayek are, to be sure, distinct thinkers with different views—sometimes fundamentally different views—on many issues in economic theory and method. There is a definite contribution to be made, towards properly understanding

each of these two great Austrian economists, by drawing attention to the matters on which they disagree. But, we must insist, (a) the general character of the market process does *not* constitute such an area of disagreement; and (b) this area of shared understanding is so central to the work of both Mises and Hayek, that our awareness of their common position in this matter must definitively dispel any suggestion of the existence of a Misesian “paradigm,” in regard to the market process, that is sharply to be distinguished from a Hayekian “paradigm.” Yet such claims have recently been made.

Mises and Hayek Dehomogenized?

Professor Salerno has, in a number of recent papers (1990, 1991, 1993, 1994), initiated a line of intellectual historiography designed to drive a wedge between Mises’s and Hayek’s understanding of markets. Murray Rothbard and Jeffrey Herbener (Rothbard 1991, 1992, 1994; Herbener 1991) have hailed Salerno’s thesis as providing definitive grounds for the rejection by all “Misesians” of what Salerno, Rothbard, and Herbener see as grave “Hayekian” errors.¹

The asserted distinctions on the basis of which Salerno declares the existence of two paradigms, a Misesian and a Hayekian, can be summarized as follows: (a) Hayek was trained under Wieser, and this accounts for his failure to have absorbed the Mengerian insights which, through the teachings of Böhm-Bawerk, later matured into the Misesian position (Salerno 1993, p. 114); (b) Hayek believed that “in order for prices to fulfill their knowledge-disseminating and plan-coordinating functions, the economy must subsist in a state of (what Salerno calls) ‘proximal equilibrium,’ wherein realized prices are always fairly accurate indicators of future prices” (p. 128); Mises, on the other hand, considered the concept of equilibrium as only a mental tool. It “is impossible to determine and meaningless to suggest that the real economy is closer to the FSR [final state of rest], and therefore manifests a superior coordination of plans and greater allocative efficiency, at one instant of time than it was at a previous instant” (p. 129). The social role fulfilled by prices does not depend on the attainment or near attainment of the FRS. This leads directly to the next point. (c) For Hayek, allocative efficiency consists in plan coordination among market participants. For Mises, on the other hand, the social efficiency achieved by the market consists (and is *always* perfectly attained) in the *ex ante* “appraisal and allocation of resources [by entrepreneurs] in strict accordance with anticipated consumer preferences” (p. 130). Salerno recognizes that, in regard

¹Because of Salerno’s initiating and prominent role in the “two-paradigm” literature, this section refers primarily to his writings. However, similar statements can typically also be found in the above cited papers of Rothbard and Herbener.

to *ex post* efficiency, entrepreneurial errors are inevitable in a world of uncertainty and change. However, apparently the only systematic process which Salerno recognizes in Mises as tending to correct such *ex post* inefficiencies, is that in which less astute entrepreneurs come to be weeded out of the system through their repeated speculative failures and resulting losses (pp. 131ff). (d) For Hayek the essence of the market process and of its social function, is in its overcoming of the "knowledge problem" arising out of dispersed knowledge "among the multitude of individual consumers and producers" (p. 115). It is this property of the market, and its absence in the socialist economy, which identified, for Hayek, the fundamental weakness of socialist planning. For Mises, on the other hand, Salerno and his colleagues claim, even if the socialist planners were miraculously endowed with perfect information, they would nonetheless be unable to "rationally calculate how to combine resources to render efficient production" (Herbener 1991, p. 43).

It is, indeed, especially the interpretation of Mises's thesis concerning the impossibility of socialist economic calculation that has been perhaps the central focus of Salerno's "two-paradigm" thesis. After a number of pages in which Salerno (quite unsuccessfully, it must surely appear) seeks to refute Leland Yeager's definitive paper (Yeager 1994) demonstrating that Mises's thesis does, after all, require that we attribute to Mises at least implicit recognition of Hayek's "knowledge problem," Salerno sums up as follows: "Thus market oriented PC [i.e., perfect competition] theorists, such as Hayek and Yeager, and neo-classical/socialist GE [i.e., general equilibrium] theorists are brothers under the skin" (Salerno 1994, p. 119).² Let us indeed, then, take up Salerno's treatment of the Misesian thesis; it will, I believe, permit us to confront Salerno's major points of contention. We shall, I further believe, be able in this way to place our finger not only on the source of the two-paradigm fallacy, but (at the same time), also on a significant element in Mises to which Salerno has properly drawn attention. The circumstances that Salerno's recognition of this element in Mises has, in our judgment,

²The biting sarcasm employed in this assertion is but a relatively mild example of the rhetorical excesses appallingly to be found in the "two-paradigm" literature against such writers as Hayek, Lachmann, and others charged with having diverged from the asserted "Misesian paradigm." I take this opportunity strongly to protest the use of verbal terrorism in Austrian economics. Even if (which is far from being the case) the asserted criticisms of Hayek, Lachmann, and others were valid, there would be absolutely no justification for the manner in which these great economists have been treated in the literature under discussion. The near-demonization of Hayek and Lachmann for alleged deviations from an asserted Misesian orthodoxy is a most distressing phenomenon. If Austrian economists (and the *Review of Austrian Economics*) are to be able to work constructively in the rough and tumble of the intellectual market place, anything approaching rhetorical brawling must once and for all be rejected.

unfortunately misled him (and Rothbard) to see *fundamental* divergence where none exists, should not blind us to the value of this characteristically Misesian insight for Austrian economic understanding.

Mises and the Calculation Problem

Salerno and Rothbard are fully justified in emphasizing the subtlety of the Misesian concept of economic calculation. With much of what they say in exposition of that concept, this writer is in full agreement. He objects only to the quite unwarranted conclusion which they draw from that exposition to the effect that the Misesian calculation problem has nothing whatever to do with Hayek's knowledge problem. A possible contribution to this unfortunate misunderstanding lies, I believe, in Hayek's earlier ambiguity concerning the nature of his knowledge problem. This writer has for a number of years (see Kirkzner [1984] 1992, p. 149), pointed out that Hayek's brilliant 1945 paper, "The Use of Knowledge in Society," was seriously confused in making it appear that the function of prices in communicating knowledge was a function that is filled, in principle, also in the state of equilibrium. Salerno and Rothbard would be on firm ground if they objected, as this writer has objected, to such an equilibrium treatment of the place of knowledge and the communication function of prices. But the truth is (as becomes evident in Hayek's later work, see especially Hayek [1968] 1978) that Hayek's knowledge problem relates fundamentally to those states of affairs in which—precisely because of the knowledge problem—market agents are making plans which do not, in the fullest sense of the term, dovetail with each other.

As Salerno and Rothbard point out, calculation is needed in order to appraise the wisdom of prospective action. Without the tool of genuine money prices, economic agents would be reduced to comparing goods sacrificed and goods received, in the face of their obvious heterogeneity and incommensurability. Such an agent would be called upon, in effect, (except in the simplest of Crusoe economies), to make decisions with his eyes closed; he would have no way of knowing whether his outcome represents profit or loss. Market prices provide the indispensable tool needed for calculation. Because the socialist society does not include resource markets, its central planners must operate without known resource values. Their decisions must be made, in effect, with eyes closed.

Under capitalism, entrepreneurs make their plans based on their entrepreneurial awareness of the resource prices they must pay in the more immediate future, and of the product prices they anticipate that they will be able to command in the more remote future. These anticipated prices provide the entrepreneur with cardinal numbers on the basis of which to appraise the profitability (or its absence) of prospective

entrepreneurial activities. In the absence of resource prices under socialism, rational central planning is literally impossible, as Mises stated (and as Salerno and Rothbard quite correctly emphasize in their interpretation of Mises).

Where Salerno and Rothbard have (as demonstrated by Yeager) gone astray,³ is in their refusal to recognize that this impossibility of rational calculation and action under socialism *can* illuminatingly be recognized as arising out of the limitations of the human planning mind—in other words, as consisting in a disastrous knowledge gap which, without market prices for resources, it is impossible to bridge. We may readily concede that Mises did not articulate his calculation problem in terms of knowledge; but this does not in the slightest imply that that problem *cannot* be seen to consist of a knowledge problem. Reasonable interpreters of Mises may disagree on whether (as this writer emphatically believes to be the case) Mises's calculation problem is indeed seen *more* clearly when its knowledge implications are made explicit. But there is no basis whatever for claiming that, in exposing these knowledge implications of the Misesian argument, one is distorting or falsifying that argument.

To be unable to calculate the worthwhileness of a prospective action taken in a market society, is, after all, *to not know the importance to others* of the goods and services one commits to that action, and the importance to others of the goods one will obtain from that action. It is quite true, that Mises pointed out (and Salerno and Rothbard cite this again and again) that the calculation problem would exist even for a socialist planning authority possessing on its desks and in its computer memories, the fullest technological information of the age, full information on available resource availabilities, and full (and somehow, unanimous) information of the social ranking of the importance of ends. This is because, even armed with such "knowledge" (or, perhaps, precisely because the authority would be engulfed by these floods of information), the members of the authority would still *not know what they would need to know*, in order to calculate. As Leland Yeager has explained, *possessing* all this information is not the same as having *assimilated* it, and having been able to deploy it (whether by computing the solution to simultaneous equation systems, or whatever) to discover the relative values of the relevant resources and products. The members of the authority

³This paper concentrates critically only upon those aspects of Salerno's and Rothbard's papers which are directly relevant to our placing the market process at the center of Mises's system. We do not take up here any criticism of a number of related assertions contained in these papers (concerning: entrepreneurship, uncertainty, the future, alertness, discovery, and coordination) which this writer finds puzzling, contradictory, or otherwise based on possible misunderstanding.

would not know what one needs to know in order to calculate the worthwhileness of prospective decisions.

For Mises (as Salerno and Rothbard correctly point out) prices are not primarily signals economizing on the cost of *communicating* information.⁴ Their social function consists in providing decision makers with meaningful cardinal numbers with which to calculate the worthwhileness of prospective actions. To be “meaningful” we do not require these cardinal numbers to be roughly equal to or close to relevant equilibrium values. We require only that, at each point in time, these cardinal numbers reflect the interplay of the decisions made by the keenest (as well as those less keen) of the entrepreneurial minds in the market economy. In all this, I am in complete agreement with Salerno and Rothbard.

But it is precisely here, I believe, that Salerno and Rothbard have, in properly drawing attention to an underemphasized element in Mises’s position on economic calculation, been led into error. The element being here referred to is that, for Mises, even market prices that are very far from their equilibrium values perform a valuable role in enabling entrepreneurs to calculate. Let me emphasize even more starkly the aspect of this element in Mises which appears to have most impressed Salerno and Rothbard: *Even if we could imagine that the equilibrating market process has not yet succeeded in nudging disequilibrium prices at all towards equilibrium, these prices yet perform their social role in making possible economic calculation.* It is apparently this aspect of the Misesian position which has taught Salerno and Rothbard that what makes calculation possible cannot be and is not that knowledge-enhancing process which, for Hayek and other Austrians, constitutes the process of market equilibration. It followed, for these two scholars, that the Misesian calculation problem under socialism cannot and must not be identified with the Hayekian knowledge problem (which tends to become solved during the course of the equilibrating market process). But there is no reason at all to arrive at such an understanding (or, rather, misunderstanding) of Mises’s position.

False Prices and Less False Prices

As cited earlier, Mises certainly did recognize that disequilibrium market prices are, in a sense, “false prices”: they reflect erroneous expectations (i.e., erroneous “knowledge”) being held by entrepreneurs

⁴This is the aspect of Hayek’s 1945 paper which the mainstream literature (and now Salerno, et al.) have seen as central to Hayek’s position. This writer has long deplored according centrality to such a “communication” role, and has argued that Hayek’s later work suggests that he, too, saw beyond such a narrow interpretation of the role of prices (see Kirzner [1984] 1952, chap. 8).

concerning the true preferences of consumers. It is the equilibrating force generated by the process of entrepreneurial competition, we saw, which for Mises tended to replace false prices by less false prices. We have every reason to believe that, when Mises sees market prices as effective tools for entrepreneurial calculation, his view of prices is, at the very least, rendered even more benign by his understanding of the market process in which *earlier* false prices have tended to have become replaced by less false prices. (Of course this tendency *may* be frustrated by entrepreneurial error in an uncertain, changing world. There is no *guarantee* that today's prices are *necessarily* less false than yesterday's. But this possibility does not eliminate the existence of a systematic process in which entrepreneurial profit-seeking activity identifies those false prices which promise pure profits, and, by grasping those profits, tends to replace them by prices which more accurately reflect the true values to consumers, of resources and products.)

Salerno and Rothbard are right to emphasize that for Mises the prices which prevail at *any* time fulfill their function of rendering economic calculation possible. This, we must insist, is *not* because all prices, at all times, are "market clearing prices," in any sense relevant for our evaluation of the social efficiency of the price system. After all, false prices reflect production plans which *are*, by definition, at variance with the true preferences of consumers. The Misesian insight that all prices, at all times, render economic calculation possible, arises out of two closely related circumstances: (a) at each instant in time, the price offers and bids, and thus also the realized prices, reflect the expectations of the most canny entrepreneurs in the market (so that what may, a day later, with the wisdom of hindsight, indeed be seen as having been false prices, were nonetheless, in terms of the most perceptive entrepreneurial assessment of the preceding day, at that time expressive of the most judicious readings—the best *knowledge*—of consumer preferences); (b) at each instant in time, current prices *are* the outcomes of processes of entrepreneurial profit-seeking corrections of still earlier false prices; at no time, in the real world, can we say that the corrective market process has not yet begun its work. At each instant, therefore, current market prices reflect the best conceivable estimates of relative consumer preferences. The calculations which entrepreneurs make by reference to such prices (and by reference to such expected future prices), are thus informed by the assessment of the shrewdest of entrepreneurs, operating under the powerful incentive of winning pure profits.

What we wish to stress is that the capacity of market prices to inspire calculative economic activity is based solidly on the extent to which prices do express correct assessments of (i.e., the relevant

knowledge regarding) both current and future preferences of consumers, and the current and future production plans of other entrepreneurs. As Mises pointed out in his first statements on the calculation problem (see, e.g., Mises [1922] 1936, pp. 115–17), market prices are not perfect tools in this respect: but they are extraordinarily valuable tools nonetheless. Their value surely lies in the expression of the best available entrepreneurial knowledge concerning market conditions.

It is quite true that for Mises this “best available entrepreneurial knowledge” expressed in current market prices would be valuably useful for calculation purposes *even if* one could imagine these prices *not* already to reflect the corrective entrepreneurial market process which tends to replace false prices with prices less false. But the circumstance that in fact current market prices reflect that corrective market process (and our awareness that Mises did indeed emphasize this circumstance in regard to market prices) should convince us that an appreciation of the role of market prices stated in terms of the “Hayekian” knowledge problem is simply a somewhat differently articulated appreciation for the calculative properties Mises taught us to understand to exist in those market prices.

Some Observations on the Misesian Legacy

Mises had a profound and subtle understanding of the market’s operation. In that understanding, the character of the market as a process in which mistaken entrepreneurial judgments tend to come to be replaced by more accurate judgments (and thus one in which false prices are replaced by less false prices), was a central feature: Hayek, too, had his own understanding of the market’s operation. In certain respects, particularly in its articulation of the role of knowledge and discovery, that understanding can be differentiated from that of Mises. *But the centrality of the knowledge-corrective character of the market process for both Mises and Hayek cannot seriously be doubted.* Whatever the differences between a Hayekian articulation of the market process and a Misesian articulation, *the centrality of the notion of the corrective process for both, is the crucially important circumstance.* It is this that should convince us that any talk of a Hayekian “paradigm” which differs fundamentally from the Misesian paradigm should be dismissed as not only reflecting a mistaken doctrinal judgment, but as reflecting a mistaken judgment with potentially catastrophic implications for the future of Austrian economics.

Austrians are a beleaguered minority in the economics profession today. One of the core doctrinal issues separating Austrian economics from the mainstream is that Austrians understand the entrepreneurial

character of the market process. We learned this from Mises. From Hayek we learned additional, complementary insights. If we wish to preserve and build upon the Misesian legacy, we must not generate confusion (both among Austrians and their opponents) by exaggerating perceived differences between Mises and Hayek, to the point where the centrally shared insights of both are dangerously obscured.

References

- Hayek, Friedrich A. von. 1978. "Competition as a Discovery Procedure." In *New Studies in Philosophy, Politics, Economics, and the History of Ideas*. Chicago: University of Chicago Press. Chap. 12. This paper was delivered as a lecture in 1968.
- Herbener, Jeffrey M. 1991. "Ludwig von Mises and the Austrian School of Economics." *Review of Austrian Economics* 5 (2): 33–50.
- Kirzner, Israel M. 1984. "Prices, the Communication of Knowledge and the Discovery Process." In Kurt R. Leube and Albert H. Zlabinger, eds., *The Political Economy of Freedom: Essays in Honor of F. A. Hayek*. Vienna: Philosophia Verlag. Reprinted in Kirzner, *The Meaning of Market Process: Essays in the Development of Modern Austrian Economics*. London and New York: Routledge, 1992. Chap. 8.
- Mises, Ludwig von. 1936. *Socialism: An Economic and Sociological Analysis*. J. Kahane, trans., London: Jonathan Cape; German original published in 1922.
- . [1951] 1962. "Profit and Loss." Reprinted in *Planning for Freedom*. 2nd ed. South Holland, Ill.: Libertarian Press. Chap. 9.
- . 1966. *Human Action: A Treatise on Economics*. 3rd rev. ed. Chicago: Contemporary Books.
- Rothbard, Murray N. 1992. "The Present State of Austrian Economics." Working Paper of the Ludwig von Mises Institute. November.
- . 1994. Review of *Austrian Economics: Tensions and New Directions*. Bruce Caldwell and Stephan Boehm, eds., *Southern Economic Journal*. October.
- . 1991. "The End of Socialism and the Calculation Debate Revisited." *Review of Austrian Economics* 5 (2): 51–76.
- Salerno, Joseph T. 1990. "Ludwig von Mises as Social Rationalist." *Review of Austrian Economics* 4: 26–54.
- . 1991. "Commentary: The Concept of Coordination in Austrian Macroeconomics." In Richard M. Ebeling, ed., *Austrian Economics, Perspectives on the Past and Prospects for the Future*. Champions of Freedom Series. Vol. 17. Hillsdale, Mich.: Hillsdale College Press.
- . 1993. "Mises and Hayek Dehomogenized." *Review of Austrian Economics* 6 (2): 113–46.
- . 1994. "Reply to Leland B. Yeager on 'Mises and Hayek on Calculation and Knowledge'." *Review of Austrian Economics* 7 (2): 111–25.
- Yeager, Leland B. 1994. "Mises and Hayek on Calculation and Knowledge." *Review of Austrian Economics* 7 (2): 93–109.

The Review of Austrian Economics, the predecessor to the *QJAE*, was founded and edited by Murray N. Rothbard and functioned as the premier Austrian School scholarly journal between 1987 and 1997. After 1995, it was edited by Walter Block, Hans-Hermann Hoppe, and Joseph T. Salerno. Published first by Lexington Books and later by Kluwer Academic Publishers, the journal played a central role in attracting a new generation of economists to the theoretical foundations established in the work of Carl Menger and Ludwig von Mises.