THE REVIEW OF AUSTRIAN ECONOMICS

Volume 9, Number 2

THE REVIEW OF AUSTRIAN ECONOMICS

Volume 9, Number 2

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

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.

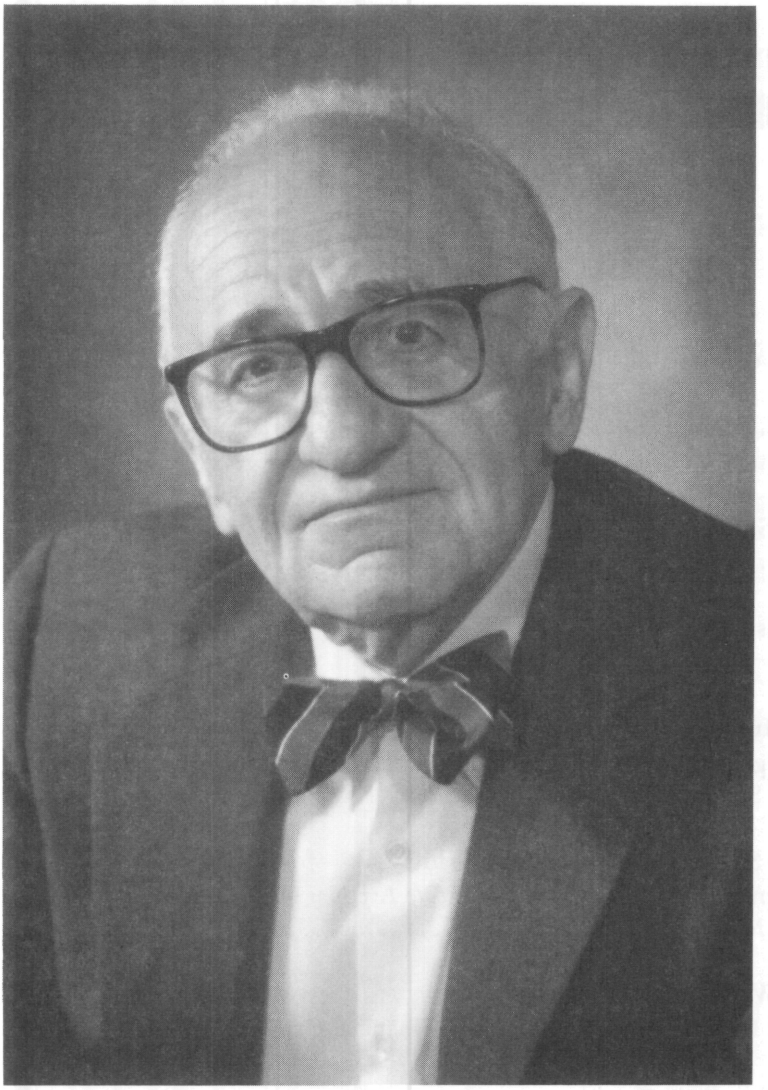
The opinions expressed in this volume by the authors in their respective articles are their own and *The Review of Austrian Economics* assumes no responsibility for these views.

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

VOLUME 9, NUMBER 2 **1996**



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

Morgan Adams, Jr.
The Adams Fund

O. P. Alford, III†

Anonymous (3)

George E. Atkinson, Jr.

Donald Bently

John Hamilton Bolstad

Philip G. Brumder

Robert H. Clarke

Douglas J. Cook
Cook Brothers Manufacturing

Mr. and Mrs. William C. Daywitt

Maino and Dorothy des Granges

Richard A. Elwyn

Dr. Larry J. Eshelman

James and Vickie Feutz

Willard Fischer

James W. Frevert

John F. Gemmill

Henry Getz

Hager Lumber Company

Frank W. Heemstra

In memory of
John A. and Erlene Hendrickson

Albert L. Hillman, Jr.

John Hotonian

Hugh P. King

Robert H. Krieble

Hugh E. Ledbetter

Winston Lehman

Arthur L. Loeb

Jack E. Magoulakis

David Manley

Dr. Walter Marcyan

Arthur Marron

Forrest E. Mars, Sr.

William W. Massey, Jr.

Roy G. Michell, Jr.
Roy G. Michell Foundation

Haydock H. Miller, Jr.

Frank and Floralie Millsaps

M. K. Morse

The Nehemiah Foundation

Mr. and Mrs. Brantley I. Newsom

Ellis L. Phillips, Jr.

Clifton B. Rodes

James M. Rodney

Catherine Dixon Roland

Sheldon Rose

E. D. Shaw, Jr.†

Quinten E. Ward

David and Barbara Westrate

Jerald N. Wigutow

Thomas G. Winter

John R. Wright

Dr. Carl E. Youngstrom

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.

[1]For 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 Review of Austrian Economics Vol. 9, No. 2 (1996): 3–28
ISSN 0889-3047

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.

[2]The 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

[3]I have in mind recent work by Boudreaux and Holcombe (1989), Foss (1993a, 1993b, 1993c), Langlois (1988, 1992a, 1992b), and Minkler (1993a, 1993b).

[4]For 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.

[5]Indeed, 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

[6]Austrians 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

[7]For 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).

[8]Other 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

[9]It 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 known[11]

[10]Lange 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).

[11]We 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,

[12]Mises 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

[13]Rothbard (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]

[14]This 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.

[15]Lavoie 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)

[16]This 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.

[17]Mises (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 Markets[18]

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

[18]This section is based on Klein (1994, pp. 397–98).

[19]There 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]

[20]The 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.

[21]Compare 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.

[22]Chapter 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

[23]On 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]

[24]Nor 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 "taxis," 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).

[25]For 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.* Aldershott, 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–Profit Center 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: Liberty*Classics*, 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 Rothbard[1] 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.

[1]Particularly 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.

The Review of Austrian Economics Vol. 9, No. 2 (1996): 29–42
ISSN 0889-3047

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.

[2]The 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.

[3]Anyhow, 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

[4]We 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

[5]Murray 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

[8]Ibid., p. 570.

[9]Moreover, 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

[10]This 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

[11]As 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

[12]K. 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

[13]Rothbard, *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 Coase[16] 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

[16]Ronald 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.

[17]Franchising 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 Fisher[5] and Edwin R.A. Seligman[6] both agreed that large-scale production produced

[1]Richard T. Ely, *Monopolies and Trusts* (New York: MacMillan, 1990), p. 162.

[2]John Bates Clark and Franklin Giddings, *Modern Distributive Processes* (Boston: Ginn & Co., 1888), p. 21.

[3]Herbert Davenport, *The Economics of Enterprise* (New York: MacMillan, 1919), p. 483.

[4]James L. Laughlin, *The Elements of Political Economy* (New York: American Book, 1902), p. 71.

[5]Irving Fisher, *Elementary Principles of Economics* (New York: MacMillan, 1912), p. 330.

[6]E. 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

[7]Ibid., p. 97.

[8]Simon Patten, "The Economic Effects of Combinations," *Age of Steel* (Jan. 5, 1889): 13.

[9]Franklin Giddings, "The Persistence of Competition," *Political Science Quarterly* (March 1887): 62.

[10]David A. Wells, *Recent Economic Changes* (New York: DeCapro Press, 1889), p. 74.

[11]George 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

[12]A. W. Coats, "The American Political Economy Club," *American Economic Review* (Sept. 1961): 621–37.

[13]Thomas 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

[14]Burton 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.

[16]George 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

[18]Ibid., p. 31.
[19]Ibid.
[20]Ibid., p. 47.
[21]Ibid., p. 52.
[22]Ibid., 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

[23]Ibid., p. 106. Emphasis added.

[24]Horace M. Gray, "The Passing of the Public Utility Concept," *Journal of Land and Public Utility Economics* (Feb. 1940): 8.

[25]Ibid.

[26]Ibid., p. 9.

[27]Ibid.

[28]Ibid., p. 15.

[29]Ibid., 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]

[30]George Stigler and Claire Friedland, "What Can Regulators Regulate? The Case of Electricity," *Journal of Law and Economics* (October 1962): 1–16.

[31]Gregg A. Jarrell, "The Demand for State Regulation of the Electric Utility Industry," *Journal of Law and Economics* (October 1978): 269–95.

[32]Demsetz, *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.

[33]Murray N. Rothbard, *Power and Market: Government and the Economy* (Kansas City: Sheed Andrews and McMeel, 1977), pp. 75–76.

[34]Murray 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,

[36]Ibid., p. 548.

[37]Edwin 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.

[38]Harold Demsetz, "Why Regulate Utilities?" *Journal of Law and Economics* (April 1968): 55–65.

[39]Ibid.

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 were 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!

[40]Ibid.

[41]Steve Hanke and Stephen J. K. Walters, "Privatization and Natural Monopoly: The Case of Waterworks," *The Privatization Review* (Spring 1987): 24–31.

[42]Walter 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.

[44]The 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,

[47]Thomas Hazlett, "Duopolistic Competition in Cable Television: Implications for Public Policy," *Yale Journal on Regulation* 7 (1990).

[48]Ibid.

[49]Ibid.

[50]Ibid.

[51]Thomas 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

[53]Thomas Hazlett, "Duopolistic Competition in Cable Television."

[54]Adam 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.

[56]Ibid.

[57]G. H. Loeb, "The Communications Act Policy Toward Competition: A Failure to Communicate," *Duke Law Journal* 1 (1978): 14.

[58]Thierer, "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.

[59]Gray, "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.

[1]See 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.

[2]Ramón 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

[3]Finally, 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.

[4]See Carlo M. Cipolla's important article "La moneda en Florencia en el Siglo XVI," published in *El Gobierno y la Moneda: Ensayos de Historia Monetaria* (Barcelona: Editorial Crítica, 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: Societá 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]

[5]Cipolla 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.

[6]Among 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 Crítica, 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).

[7]Luis Saravia de la Calle, *Instrucción de Mercaderes* (Medina del Campo: Pedro de Castro, 1544); republished in *Colección de Joyas Bibliográficas* (Madrid, 1949).

[8]Ibid., p. 180.

[9]Ibid., 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

[10]Ibid., p. 195.

[11]Ibid., p. 196.

[12]Ibid., p. 197.

[13]Ibid.

[14]See 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

[15]Saravia 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* (*Resolutory 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

[18]Martín 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."

[19]Ibid., pp. 60–61.

[20]Ibid., p. 61.

[21]I 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

[22]Tomás 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

[24]This 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.

[25]According 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

[26]It 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.

[27]Luis de Molina, *Tratado sobre los Cambios*, edition and introduction by Francisco Gómez Camacho (Madrid: Instituto de Estudios Fiscales, 1990), pp. 137–40.

[28]Ibid., 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.

[29]Ibid., p. 137.

[30]Ibid., 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.

[31]R.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.

[32]See, 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.

[33]The 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 expenduntur, publica, et Regia authoritate 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 Volkswirthschaftslehre* (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.

[34]Martín de Azpilcueta, *Comentario Resolutorio de Cambios*, pp. 74–75; emphasis added.

[35]Not 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

[36]Luis 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]

[37]Ibid., p. 146.

[38]See 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.

[39]Luis de Molina, *Tratado sobre los Cambios*, p. 147

[40]Ibid., 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.

[42]Ibid., pp. 406–7.

[43]That 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

[44]Published 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

[46]Ibid., p. 87. The reference to Juan de Lugo corresponds to *Disputationum de Iustitia et Iure, Tomus Secundus* pp. 60–62.

[47]Bernard 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.

[48]Father 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.

[51] This 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.

[1]With 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.

[2]This essay is reprinted in White (1989), a book Hoppe cites.

[3]Selgin'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.

[4]Perhaps 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 *ir*redeemable 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.

[5]Redeemable 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).

[6]To 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

[7]A 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.

[8]Whether 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 *un*suited 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

[9]Hoppe 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.

[10]Checkland (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 payoffs 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

[11]It 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

[12]Economists 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.

[13]In 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 noteholders, 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

[14]Likewise 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.

[15]Given 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."

[16]It 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]

[17]Some 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.

[18]The 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

[19]Picture 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*."

[20]It might be said that most people would rather "put their money where the moths are."

[21]For 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.

[22]Rothbard (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.

[23]In 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 Commision 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

[24]Our 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.

[25]Hoppe (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 *in*consistency, 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

[26]Thus 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

[27]Thus 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

[28]It 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.

[29]But 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 Disequlibrium." *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

[1]The 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 provid 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 exposited 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

[2]It 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.

[3]Here 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.

[4]Even 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]

[5]By wholly ignoring discoordinating 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 qs, is accompanied by a decrease in the corresponding ps. It would be misleading here to evoke the fears of "deflationary pressures." The individual ps become adjusted to their corresponding qs on a market-by-market basis. The fact that this new constellation of ps 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

[6]Selgin (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 exposited 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

[7]The 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 deceivingly 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

[8]There 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

[9]This 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

[10]The 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 ps 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.

[1]For a survey of this literature, along with some effort to reconcile the various strands, see Vaughn and Wagner (1992).

[2]For a good sample of this controversy between these two positions, see the essays in Ferguson, ed. (1964).

The Review of Austrian Economics Vol. 9, No. 2 (1996): 129–42
ISSN 0889-3047

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 behaviorial 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

[3]The 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 marshall 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

[4]On 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 plausiblity 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

[5]Or to cover reductions in tax receipts below normal levels in the case of recessions.

[6]This 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 reponse 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

[7]See, 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

[8]This 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.

[9]Wicksell 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.

[10]See, 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.

[11]See, 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.

[12]The churning state is articulated nicely in de Jasay (1985).

[13]A 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 transations. 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* "appraisement and allocation of resources [by entrepreneurs] in strict accordance with anticipated consumer preferences" (p. 130). Salerno recognizes that, in regard

[1]Because 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 neoclassical/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,

[2]The 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 Kirzner [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 knowlege 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

[3]This 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

[4]This 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.

JOURNAL DES ECONOMISTES ET DES ETUDES HUMAINES

A bilingual journal of interdisciplinary studies

Volume VI Numéro 1 Mars 1995

ARTICLES

COMMUNICATIONS ET COMMENTAIRES

COMPTES RENDUS ET NOTES DE LECTURE

Annual (3 issues) France subscriptions: FF200 students with copy of ID, FF300 individuals, FF450 institutions. Outside France: FF250 student or US$45, FF350 individuals or $75, FF500 institutions or $90.
Send check order to:
IEEH, Centre d'Analyse Economique, 3 Av. Robert Schuman, 13628 Aix en Provence France
tel: (33) 42 96 54 20; fax (33) 42 96 96 30)
or transfer the amount to:
BPPC, Account# 6019263955, C/Bank# 14607, C/guichet #00060, C/Rib #67 Aix République
Account Name: Journal des Economistes

THE JOURNAL OF LIBERTARIAN STUDIES

VOLUME 12, NUMBER 1, SPRING 1996

Dedicated to the Memory of Murray N. Rothbard

THE JOURNAL OF LIBERTARIAN STUDIES is published semi-annually by the Center for Libertarian Studies, a nonprofit, privately supported institute dedicated to research and scholarship and the ideal of a free society. Annual subscription rates are $40 for institutions, $30 for individuals, and $20 for students. Canada and Mexico subscribers please add postage as follows: for surface delivery $4, for airmail $6. All other international subscribers please add $10 (U.S. dollars only, please, Visa and Mastercard accepted). Address inquiries to THE JOURNAL OF LIBERTARIAN STUDIES, P.O. Box 4091, Burlingame, California, USA 94011, Telephone (415) 692-8456 or (800) 325-7252. ISSN 0363-2873