# THE ARNE RYDE SYMPOSIUM ON THE ECONOMIC THEORY OF INSTITUTIONS

Editors: Lars Jonung and Ingemar Ståhl

# THE ARNE RYDE SYMPOSIUM ON THE ECONOMIC THEORY OF INSTITUTIONS

# Introduction

The purpose of this special issue of the *Statsvetenskaplig Tidskrift* is to present contributions in economics which are closely related to the subject matter of other social sciences. The papers printed are a sample of the contributions to the symposium on the Economic Theory of Institutions held at Frostavallen outside Lund in September 1979. This symposium was sponsored by the Arne Ryde Memorial Fund and the administrative responsibility was undertaken by the Department of Economics of the University of Lund. The scope of the symposium was to discuss the behaviour of, and the interactions among, institutions in a society of the mixed-economy type. The main areas of interest were property rights and contracts, public choice and institutions such as unions, bureaucracies and different types of enterprises.

It is fitting that these papers should be published by the *Statsvetenskaplig Tidskrift* (The Swedish Journal of Political Science). For a long time the *Statsvetenskaplig Tidskrift* has contained contributions not only in political science but also in economics. When neo-classsical micro-theory and Keynesian macro-theory dominated economic science, Johan Åkerman, Professor of Economics in Lund between 1944 and 1961 argued strongly for a more synthetic approach to research in economics in which the institutional framework and political behaviour are an integral part. Åkerman was a pioneer in studies of the interaction between economic and political developments. From his many contributions in this area one cannot escape the feeling that many-problems that now seem new and fresh were actually part of earlier economic science and that the interaction between economics and politics has been sadly neglected over the past couple of decades.

Economic science in Lund has a long, although interrupted, tradition in being integrated with other social sciences, especially political science.

Erik Lindahl's thesis *Die Gerechtigkeit der Besteuerung* was admitted in Lund in 1919. He was obviously building on the work of one of the giants of economic science, Knut Wicksell, who was Professor of Economics and Fiscal Law in Lund between 1900 and 1916. Wicksell's seminal work *Finanztheoretische Untersuchungen* in 1896 marks the beginning of much of what is now called public choice. Not only did he formulate the unanimity principle for the determination of optimal allocation of public goods but he also studied the implications of different voting rules.

We can, perhaps stretching the point a little, trace the tradition further back through the centuries. Just after the University of Lund was founded in 1668, the German scholar Samuel von Pufendorf was appointed Professor in Law and Jurisprudence. From his work *De officio hominis et civis juta legem naturalem,* first published in Lund in 1673, we quote the introduction to his chapter on price formation:

"After ownership had been introduced, and since all things were not of the same nature, and did not yield the same services to human necessities, and no one had that abundance which he desired for his needs, it soon became customary among men to exchange commodities." (English translation from the Latin original in the edition published in New York in 1927.)

Appropriately, this volume begins with a paper on property rights in which Carl J. Dahlman attempts to establish the nature of the connection between property rights and economic institutions. The second paper by Mats Lundahl analyses how the system of property rights in Haitian agriculture has been influenced by population growth.

The next two papers by James M. Buchanan and Dennis C. Mueller develop aspects of organizations that are largely neglected in traditional micro-theory. Buchanan discusses the role of en-

trepreneurship and Mueller develops an analogy between power and profits as driving forces.

The contributions by Svetozar Pejovich and Per-Olof Bjuggren each provide examples of enterprises that differ from the conventional neoclassical "capitalist" firms with regard to ownership and incentive structure. Both authors treat some important aspects of the behaviour of labour-managed firms with special reference to firms of the Yugoslav type. Bjuggren's paper also contains an empirical comparison between firms of a traditional type owned by private stockholders and firms owned by producer cooperatives.

Levis A. Kochin estimates the social costs of unions due to the mis-allocation of labour between unionized sectors with relatively high wages and non-unionized sectors with lower wages and to different types of rent-seeking behaviour.

The last three papers consider different features of government. Michael D. Bordo and Daniel Landau interpret government as being essentially a "protection agency" for the lives, property and permanent income of its citizens. Friedrich Schneider and Werner W. Pommerehne discuss the issue of fiscal illusion. Finally, Thomas Wilson examines the role of economists and welfare theory with respect to the welfare state.

This collection of papers is not intended to provide comprehensive coverage of the economic theory of institutions. Rather, the aim is to show how new thinking in economics might facilitate a better understanding of some areas of mutual interest to economists and other social scientists.

Lund, February 1981

*Lars Jonung*                    *Ingemar Ståhl*

# On Property Rights, Transaction Costs, and Economic Institutions

BY CARL J DAHLMAN*

## 1. Introduction

What I aim to do in this paper is to make some rather simple observations on the efficiency of economic institutions, and to approach that from an oblique historical angle. When it comes to economic institutions I think we all, of sheer necessity fostered by the nature of the problem, must become historians, at least if we have any interest whatsoever in empirical application and falsification. An historical perspective is necessary since institutions, of all the endogenous choice variables in an economic system, seem to be the slowest to change. If we believe that the implementation of institutions in society is due to a describable, rational choice process, then the implication is that institutions ought to change in a predictable fashion when some specifiable exogenous conditions change. To test such propositions necessarily implies doing economic history, since institutions change so slowly. If this were not an established and accepted fact, how else can we explain the traditional treatment of institutions in economics – as exogenously given, by historical evolution or otherwise, and "setting the stage", as it were, for whatever else we wish to study, i.e., notably trading arrangements.

However, there is another reason why an historical approach should prove fruitful. Suppose we do accept, and some may not, that institutions are endogenous choice variables. If we were to apply standard economic analysis, we might guess that, as long as we specify the constraints properly, the unavoidable implication will be that an optimal choice exists. The historical implication would then be that if two economic societies had similar constraints, they should also have similar institutions. I think any randomly picked economic historian would balk at such a preposterous suggestion. But if we seriously wish to apply choice theory to the issue of endogenous economic institutions, how can we possibly avoid this implication?

I have chosen to apply the emerging property rights paradigm to the problem of economic institutions for several reasons. The first is that I believe that the notion of property rights is sufficiently general to encompass practically any economic problems. Indeed, I would go so far as to embrace completely Alchian's definition of economics as "the study of property rights".[1] I would only add as a clarification the idea originally brought forward by Commons that exchange is exchange of property rights, not of physical entities. That is to say, when economists talk about "goods", I suggest that what they really have in mind is not a commodity, but a bundle of decision making rights. It is misleading to think of trade as the exchange of commodities; it is better to talk about the property rights that are exchanging hands. Since institutions are intimately connected with property rights, I therefore believe that it may be fruitful to put the creation of economic institutions into an exchange paradigm where rights or decision making powers are the objects of exchange.

A second reason for employing the property rights paradigm is that the property rights literature, more than any other branch of contemporary economics, focuses on the implications of the ubiquitous existence of transaction costs. It is my firm belief, one that I will attempt to substantiate in the following, that transaction costs are the key to an understanding of the precise function of economic institutions as well. Naturally, this implies a specific definition of what is to be understood by the notion of transaction costs, and I will attempt to justify a particular definition below. In addition, there are some important theoretical results in property rights theory, perhaps most notably the Coase theorem, that

may or may not be applicable to a choice theory of economic institutions.

So this sets the scope for what I shall discuss in the following. I will try to establish the connection between property rights and economic institutions. I will also try to justify the economic function of economic institutions as that of dealing with transaction costs, on a special and, I believe, new and more reasonable definition of what is to be understood by the notion of transactions costs. These tasks are less of the character of formal model building than definitional; however, I hope that solving the problems of classification and definition will have some immediate implications for the way the economics discipline can and should approach the problem of endogenous economic institutions. The remainder of the paper will be concerned with some observations on these implications.

## II

One of the first, and principal, items to settle when constructing a theory of economic institutions as endogenous choice variables of an economic system must be to determine what is to be meant by the phrase "economic institutions" i.e., we must start by defining properly what will be the object of our study. I know of no widely accepted definitions of what is to be understood by "an institution" in contemporary economics. It seems to be a very loose term used to cover a wide variety of phenomena. To exemplify, we think of democracy and dictatorship as institutions, i.e., we often describe differences in political structures between different societies as the result of different institutions. We spend much time analyzing the economic implications of private and collective ownership rights as different institutions. The capitalist firm and other organizations for production are usually referred to as institutions. We often characterize monetary exchange as an institution. We also refer to ethical codes, social mores, and certain cultural behavioral patterns as institutions. Given this usage of the term, the question inevitably arises whether all these different "institutions" really have some common characteristic that would allow us to classify them as belonging to the same proper set. This question can obviously not be answered until we define the common properties of the members of the set, and then check if the items on the list just presented all have that same property.

On a very elementary level, it is clear that the

institutions just referred to above all have in common that they "set the conditions for exchange and production" rather than inherently being exchange and production activities. If we were to make this the generic characteristic of institutions, however, it would be an incomplete and rather fuzzy definition, for resource endowments and productive technology are also conditions that limit and determine exchange and production. If so, there would seem to be no particular basis for making a fundamental distinction between institutions and other constraints on economic activity. Yet, if we seriously believe that the study of institutions is a line of inquiry fruitful in itself, there must be a fundamental distinction between institutions and other constraints on production and exchange.

We might begin by considering what it is that institutions do, i.e., what their economic function is. To put this in perspective, it is useful to underline the almost complete lack of institutions in comtemporary economic theory. The dominant paradigm of modern economics, Walrasian general equilibrium systems, is almost totally void of institutional structures. Conceptually, the paradigm proceeds as follows: take as given the quantity and quality of productive resource endowments, the behavior of individual agents (utility, wealth, or profit maximization with a known objective function), and the existing technology of production as inherited from the past. With technology and endowments exogenously given, quantities demanded and supplied, along with their relative prices, are determined with simple optimization procedures, i.e., by maximizing the known objective functions subject to the known constraints. However, except for the statement that ownership of the initial endowments is well known and uncontroversially distributed among the economic agents, nothing is said about the institutional content of the conceptual framework. All institutions, except private ownership of the initial endowments, are either non-existent or simply asserted to exist in standard general equilibrium theory. The only institutional agent present is the firm, but there is no justification provided for its existence. To avoid the embarrassing implication of an indeterminate firm size, an assumption is sometimes inserted that there exists a factor of production, entrepreneurial capacity, which has no opportunity cost, exists in abundant supply, and which is necessary for the operation of a firm. However, there is really nothing in this conceptual

framework that can enable us to make a funda-
mental distinction between producers and con-
sumers. Any consumer, endowed with entrepre-
neurial capacity, can become a firm by hiring labor
and capital, but there is nothing to justify the
existence of firms per se, for firms have no charac-
teristics other than that of being producers. Hence,
firms in general equilibrium theory are not
"institutions" in any relevant sense of that word,
but only a label for anyone who produces things
for trade. That is to say, firms arise in general
equilibrium theory simply by assumption, and
there is nothing apart from this artificial assump-
tion to justify the presence of firms. If we therefore
were to simply drop the word "firm" from general
equilibrium theory and simply use the division
of agents into producers and consumers instead,
nothing whatsoever would be lost.

The modern analysis of the reasons for the
emergence and persistence of the firm as an or-
ganization centers on the ability of the firm to
decrease transaction costs. In his now classic pa-
per, Coase discussed how received price theory
ought to predict that all transactions, including
those we have come to associate with activities
of the firm, ought to occur across markets, and
that the implication is that we can explain non-
market activities, such as the orders given by the
firm to its employees, by invoking costs of using
the market mechanism.[2] Alchian-Demsetz have
further specified the transaction costs relevant for
the emergence and persisting efficiency of the firm
as those associated with the monitoring of team
production.[3] If we accept this basic approach, and
generalize its principal results, we shall arrive at
the proposition that the purpose of economic in-
stitutions is to reduce transaction costs, i.e., the
costs of organizing and completing economic ex-
changes. Incidentally, the acceptance of this pro-
position provides us with a logically pleasing ra-
tionale for the lack of institutions in standard ge-
neral equilibrium analysis: general equilibrium
models are typically set up so as to contain ab-
solutely zero costs of transaction – no matter what
definition of transaction costs we employ. Hence
they should also be void of institutions, including
firms, and this explains why firms can be brought
in only by assumption, rather than being derived
within the framework itself.

However, even if we do accept the basic pro-
position that the function of economic institutions
is to minimize transaction costs, we have in no
way solved our problem, for the phrase "trans-
action costs" is, in my opinion, one of the fuzziest

in contemporary economics terminology. Else-
where I have offered a critical analysis of this con-
cept, and I have endeavored to show that the two
most frequently employed definitions of the trans-
action costs concept really do not add any sig-
nificant new insights into the nature of the costs
associated with the exchange of goods and ser-
vices.[4] The two notions referred to are the so-
called set-up and transfer costs often employed in
the mathematically oriented literature, especially
on monetary theory. The fundamental inadequacy
of these notions appears to me to be that they
are both part of exogenously given constraints,
rather than variables over which economic
agents can exercise a measure of control by mak-
ing choices. The set-up cost is usually conceived
of as a fixed cost of making an exchange, inde-
pendent of the value or the nature of the ensuing
transaction, and the transfer cost seems little more
than a regular transportation cost under a new
name. Both are usually assumed to be associated
with the commodity to be traded, i.e., they are
indexed over commodities or markets, and their
values are assumed known and fixed. Naturally,
such assumptions are useful to make the problem
tractable mathematically, but it is not clear at all
that we learn anything new about the exchange
process from this simplistic analysis.

My conclusion has therefore been that, for the
notion of transaction cost to add a truly new ele-
ment into economic theory it must be associated
with two crucial aspects of the exchange process
that are often disregarded: the cost arising from
individual behavior, and the fact that such costs
usually are uncertain. That is to say, I suggest
that if transaction costs deal solely with the tech-
nical aspects of commodities or transportation, it
would be preferable to regard them as constraints
imposed by technology, rather than as associated
with choices over transaction activities. On the
other hand, it would seem that the costs associated
with the uncertainties of individual behavior, and
the methods available for influencing those costs,
have hitherto not received their due consideration
in the economics literature. In three other contexts
I have shown how what I have called individual-
specific transaction costs can explain various
phenomena that otherwise can be analyzed only
incompletely. In a discussion of what class of
transaction costs is consistent with the genera-
tion and persistence of externalities, I have shown
that individual-specific transaction costs is
the only class even possibly consistent with tra-
ditional interpretations of the inoptimalities as-

sociated with externalities.[5] In a model of the institutional arrangements of the English open field system, I have shown specifically how the notion of individual-specific transaction costs can explain why certain institutions existed and persisted for roughly a millennium.[6] In a discussion of the transaction costs conditions that generate the use of money as a medium of exchange, I have also shown that it is individual-specific transaction costs, rather than the trivial set-up and transfer costs, that are minimized by the abandonment of barter.[7]

Nor have I been the only one to employ the concept in recent literature, although others have given it a different name. What I call individual-specific transaction costs is in practice identical to what Williamson has called opportunistic behavior,[8] and, in a more limited manner, similar to the more restricted notion of post-contractual opportunistic behavior referred to by Alchian-Crawford-Klein.[9] However, even before these various names were applied, the fundamental notion had already been employed implicitly in the analysis of the functioning of some specific economic institutions. Notably, Demsetz showed in his analysis of why private property rights are sometimes more efficient than collective rights that private rights reduce the probability that some individuals will impose costs on others by overconsuming and underinvesting in a scarce productive resource, thus dealing efficiently with opportunistic behavior by some, i.e., with individual-specific transaction costs.[10] In their analysis of the firm already referred to, Alchian and Demsetz show how the firm can deal with the costs imposed on other team members by certain negative behavior displayed by some individuals – an exellent example of how an organization is designed to deal with opportunistic behavior or individual-specific transaction costs.

The common element in all these illustrations is the basic realization that all individuals do not behave identically even when faced with the same constraints. In modern economics it is virtually unheard of to attempt to explain various phenomena observed in the real world by invoking differences in tastes or in utility functions, the problem being that such propositions are rarely falsifiable by empirical data, since any observed differences in behavior usually can be explained by simply saying that "people are different". It is usually only in the analysis of uncertainty, where differences in the attitude towards risk between individuals play an important role, that the implic-

ations of individual behavior are explored. However, I want to propose that this may turn out to be a very fruitful avenue to pursue in the analysis of the economic function of institutions as well. The purpose of the exercise is not to explain the behavior of any particular individual or group of individuals, but only to see if the assumption of different behavior within a certain group of people can be used to explain the existence of various institutional arrangements as mechanisms for dealing with the costs associated with such differences in behavior.

Let us therefore assume the existence of a distribution of individual behavior along a scale measuring the willingness of individuals to cooperate with other economic agents in society. The shape of the distribution is not important for the present, rather general, purposes. All we need to assume is that certain people tend to be more helpful and charitable than others, for whatever reasons inherent in their personal make-up. The implication is then that certain other people are not so cooperative. Let us assume that all individuals are thus distributed along a continuum from zero to one, with zero implying an absolute unwillingness to behave favorably towards others, and one an absolute, undeviating willingness to do so. For example, some people will remain thieves, or free riders on a public good, no matter what is said or done to them, and we cannot explain their persistent deviant behavior totally by invoking differences in observable variables such as income, education, upbringing, social environment, etc. Others will never steal, no matter what their observable economic constraints may be. Now, it is clear that the existence of such differences in individual behavior has important consequences for the benefit accruing to other members of society from their economic activities as well as for the functioning of the economic system as a whole. To continue the thief example, a society which has a large proportion of natural thieves will suffer economically relative to one that has a smaller proportion: to avoid being the victim of theft, other economic agents will have to invest more economic resources in protecting themselves – i.e., thieves can impose costs on the rest of society by their behavior. On the other hand, a society with no thieves will save on those resources, and with otherwise similar resource endowments and technology thus achieve a higher level of consumption than one with thieves. The response mechanism in the society with thieves may then be one of two choices, at least: either it may let

the victims of theft deal with it as they please, or the victims may find it cheaper to organize social codes, and enforce them, that deal with thievery collectively rather than individually.

We may thus conceive of the economic function of institutions as that of dealing with such costs imposed on the rest of society by the negative behavior displayed by certain individuals in one tail of the distribution referred to above. This is the collective solution just referred to. Socially enforced rules against thievery constitute an institutional arrangement that changes the incentives for individuals to display behavior associated with the left tail of the distribution along the measure of cooperativeness. It is not just thievery, of course, that is relevant in this context. There are many other activities, not all as easily condemned as thievery, that impose costs on the rest of society. It may be the unwillingness to live up to contractual obligations undertaken in normal exchanges, for example, by delivering goods or services that do not measure up to implied or explicit conditions in various agreements, or in the non-payment of services or goods received. An appropriate generic term for this kind of activity might be "rent-seeking" behavior. The more frequent such behavior is, the greater the amount of resources that other members of society will have to devote to controlling negative behavior. However, the point is also that it is difficult, i.e., costly, to ascertain exactly what individuals in society are contained in the "bad" tail of the distribution. If it were known exactly what the probability of a particular individual behaving in an unwanted fashion is, the cost of dealing with that individual could be avoided if all transactors simply refused to deal with that difficult individual or did so only at higher prices. Thus, it is frequent that an economic agent with a bad credit record finds it more difficult, or at least more expensive, to acquire loans. This is also a common solution in insurance contracts, where the underwriter can refuse to insure certain individuals with characterisitics that would make them relatively prone to being bad risks, or charges higher premia to those more likely to end up in trouble than others. However, such knowledge about the special characteristics about an individual trader would seem to be the exception, rather than the norm. Casual empiricism tells me that lack of information about trading partners is more common in the world as we know it. Consequently, there is often an element of considerable uncertainty involved in predicting what precise individuals will be imposing costs on the rest of society. When this is the case, a society may, by social consensus, impose restrictions on individual behavior so as to either completely eliminate unwanted behavior in the tail of the distribution with rent-seeking, negative behavior or at least to decrease the incentives for individuals to display such behavior. Thus, the economic function of institutions may be conceived as serving as a mechanism for dealing with the costs of individual-specific transaction costs when the probability of each individual behaving in a certain manner is unknown, or where it would be costly to attain that knowledge.

I believe that this gives a clue to what we ultimately mean by "an institution". Formally, I should like to offer the following definition of what we should understand by the word "institutions" in economics: an institutional arrangement is a specific set of attenuated decision making rights set up for the purpose of eliminating or reducing certain unwanted individual behavior when it is costly to ascertain exactly which individuals will by their behavior impose costs on others. As will be seen, such institutions can be designed or created either through social consensus through some political decision making process, or by voluntary agreement across markets. A few illustrations may be in order.

In the beginning of this section I offered a short list of various phenomena that economists often refer to as institutions. According to the views just presented here, are these phenomena really institutions? Democracy and dictatorship are but two examples of political structures that fit the definition well. Both are characterized by different limitations on political decision making rights, and thus constitute simply differently attenuated political property rights. Private and collective ownership rights are different institutions in that they allow and disallow various individual actions, i.e., attenuate decision making rights in different ways. Monetary exchange, on the other hand, cannot be properly regarded as an institution. We do not require, except in special instances and for special reasons such as the payments of taxes and other dealings with the government, that payment be made in terms of money. Hence monetary exchange is not a limitation on decision making or property rights: should an individual wish to continue to truck and barter, he is free to do so in any society that I am acquainted with (although there are some commodities he may not be allowed to sell freely). Mo-

netary exchange is in my opinion better described as a superior payments technology, but it is not an institution, even though the transaction costs that money serves to minimize also are individual-specific. However, ethical codes, social mores, and cultural behavior patterns are most definitely institutions. They are designed to limit and circumscribe individual behavior by reducing the freedom of decision making, i.e., by attenuating property rights. Their function is to make individual behavior more predictable, and to eliminate or reduce certain unwanted or undesirable behavior by agents in the social framework. The capitalist firm is also an institution, for the firm constitutes a specific set of attenuated decision making rights over scarce economic assets. The firm is empowered to make a certain set of limited decisions over the allocation of both capital and labor, and it is empowered to do so by the original owners of these productive resources. It serves to make individual behavior more predictable, for example, by reducing the problem of shirking, thereby eliminating the problem of opportunistic bargaining behavior by the owners of large shares of the capital stock.[12]

This view of what an economic institution is and its role in society implicitly relies heavily on the notion of property rights. In the modern property rights literature, there is much discussion of the economic incentives offered by various kinds of property rights and of their efficiency characteristics. The stress is on the allocative effects of various property rights arrangements. What I wish to add to this is simply to stress the obvious but sometimes forgotten observation that property rights not only change incentives, but have a much stronger role to play, in at least two ways. First, property rights, or socially sanctioned decision making rights, serve as social control mechanisms as well. We impose property rights in order to change and control individual behavior; not necessarily the behavior of everyone in society, but perhaps only that of a certain subgroup. Secondly, unlike relative prices, property rights do not change incentives alone, but also affect income distribution. By allowing or disallowing certain actions by certain individuals, society can, by changing property or decision making rights, also affect who will have the right to the income from certain specific economic assets. However, I thank it is too limited to think of property rights simply as the relationship between human agents and commodities. I think it is fruitful to include in the notion of property rights more subtle

but equally important mechanisms for social control. I have already suggested that ethical codes and social mores be included in our definition of institutions, by virtue of the limitations on individual decision making that they imply. The point about such features of any social environment is that they make us behave in certain ways and avoid behaving in others, and therefore limit our decision making freedom over our actions, or, in other words, attenuate our property rights in making economic decisions. This is a very general interpretation of the concept of property rights: with the phrase I understand simply any decision making rights over scarce economic assets, including the personal behavior of an economic agent. On this interpretation property rights are not an alternative enforcement mechanism to, for example, religion or inherited cultural values, for both these latter examples also work through changing property rights. That is to say, my freedom of choice, and the controls society extends over my person, are not emanating solely from legislative acts, but from all sources in society that affect my freedom of choice as an individual. On this interpretation, any infringement on my decisions is a change in property rights, and a change in the institutional environment.

Property-rights theory is, therefore, an extremely general and flexible language for analyzing the role of institutions in an economic environment. We can interpret any changes in decision making rights as changes in property rights for the purposes of controlling the behavior of individuals so as to reduce the problems with adverse behavior by some that results in the imposition of costs on others. It is by changing property rights, i.e., the socially recognized and sanctioned rights to undertake certain actions or display certain behavior, that any society will effect a reduction of the number of people in the tail of the distribution where adverse behavior lies.

In the next section, I shall proceed to draw some conclusions of this view of institutions and the role of property rights. However, to sum up, what I have tried to do so far is to establish the following propositions. First, that institutions serve the basic functions of dealing with transaction costs. Secondly, that transactions costs are those real resource costs imposed on members of society by adverse behavior displayed by a subset of the members in one tail of the distribution over cooperativeness. Thirdly, that transaction costs are individual-specific and associated with uncertainty as to who will be the source of such costs, and

that individually organized solutions therefore might prove inadequate. Fourthly, that the way institutions deal with such transaction costs is to change the incentives for adverse behavior, and thus to change the distribution. Fifthly, that institutions constitute decision-making or property rights, and that property rights and institutions are social control mechanisms of a rather general kind.

## III. Some implications, observations, and extensions

If I have accomplished anything so far in this paper, it is really nothing but a couple of definitions – of transaction costs, and of institutions – and a rather intuitive discussion of how property rights, institutions, and transaction costs relate to each other. However, I believe that from this simple-minded organization of terms and concepts will follow some rather important conclusions that will have a bearing on the issue I set out to discuss in this paper – the question of whether there will exist an efficient set of economic institutions in a society with given constraints on technology and resource endowments. I propose to approach this question by simply sketching the outlines of a theory of the choice of endogenous economic institutions, given the notions of transaction costs and institutions I have suggested in the preceding.

The first point to note is that, on my interpretation of what the notion of transaction costs really conveys, the level of transaction costs is an endogenous variable for the economic system as a whole – unlike the more common definitions of the concept referred to above. By choosing different institutions, i.e., different rules and attenuated decision making rights, society can affect the distribution of individuals along the cooperation continuum by reducing the incentives for undertaking certain actions, and by improving the incentives for others. Thus, it can also affect the negative interaction between individuals in such a way as to minimize the costs that "bad" behavior imposes on more cooperative members of society. The way a society decides on its appropriate level of transaction costs is by devising institutions, and it is by controlling the endogenous choice variable institutions that society can affect transaction costs emanating from individual behavior.

We might then ask the question whether it is possible to amend the standard Walrasian approach to incorporate this fact. If we agree that institutions result out of a social consensus as to what individual behavior ought to be restricted, then it would seem that we can treat economic institutions as the result of trades or exchanges in a market-like setting. Again, the firm is an excellent example. A firm or a corporation is constituted by the acquisition of a collection of conditional decision making rights over scarce capital. That is to say, a firm or a corporation is founded when certain people voluntarily transfer the rights to use their capital to a newly created economic agent – a fictitious, juridical person with decision making powers in its own right. In exchange for their voluntary relinquishing of certain decision-making rights over scarce capital assets, the stockholders or firm owners receive an increased economic benefit, sometimes called a profit. Thus, the relationship that we call the firm, an economic institution, can be viewed as the result of an exchange or a trade between the original owners of capital, and a fictitiously created economic agent, and all for the benefit of those who trade away their decision making rights. Similarly, when laborers contract with a firm, they voluntarily agree that the firm will be empowered to make certain well specified decisions regarding the use of their scarce labor services, but the decision making rights so acquired by the firm are often severely restricted. The reason it is in the interest of laborers to agree to this limitation of their own decision making rights over their labor services is naturally that they are able to get an increased real wage as a result. However, the point is simply that we may describe the rights of decision making, or the allocation of property rights, within the firm as the result of traders or exchanges. We can, if we wish, generalize this view of how economic institutions come about, and shall then end up with what I believe is called the contractarian view of economic institutions, i.e., we may wish to view *any* economic or political institution as the result of an exchange of decision making rights within society. The Declaration of Independence puts it vividly:

"We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable rights, that among these are life, liberty, and the pursuit of happiness – that to secure these rights, governments are instituted among men, deriving their just powers from the consent of the governed, that whenever any form of government becomes destructive of these ends, it is the right of the people

to alter or to abolish it, and to institute new government, laying its foundations on such principles, and organizing its power in such form, as to them shall seem most likely to effect their safety and happiness."

What I find so interesting about this sentence is not its explicit condoning of revolution as a means of social change but the words "to secure these rights, governments are instituted among men, deriving their just powers from the consent of the governed". I read two things into these words. First, a rather clear recognition of the existence and implications of the individual-specific transaction costs, and the ability of institutions to deal with them. In order to protect a certain social environment from threats by those who, for their own personal benefit might see fit to impose costs on others, institutional rules are agreed on that empower the government to protect the rest of society from such individuals. Secondly, the explicit statement that even the government of a society is to be interpreted as the result of an economic exchange. Individuals give up certain decision making rights, i.e., they declare their willingness to abide by decisions taken by their government and thus *not* make certain individual decisions, in exchange for which they get a better kind of society, and it is precisely this which constitutes the power of the government to make certain decisions. It is true in the firm that the right to make decisions vested in the management derives from the consent of the governed, i.e., the capital and labor owners who agree to abide by certain decisions, and it is equally true that we may look at political governance structures in a similar way.

Suppose that we draw the implications of such a contractarian view of the world, i.e., accept the proposition that institutions result from economic exchanges, can we then amend the Walrasian paradigm to include economic institutions? Specifically, we may inquire whether it is possible to reinterpret a subset of the equations in the Walrasian system to represent trading, not in goods, but in decision making or property rights for the purposes of creating a set of institutions conducive to economic efficiency – or, to phrase the question differently, could it really be that the Walrasian system, popular belief notwithstanding, tacitly has been describing a complete set of institutions as well as trades in real commodities all along? For three reasons I believe that the answer is negative, i.e., the Walrasian paradigm is absolutely void of institutions, economic or otherwise.

The first has to do with the nature of exchange itself. In the contractarian view of the world, the individual is thought to voluntarily relinquish decision-making rights in exchange for a well specified economic benefit. This naturally involves the establishing of a contract – a kind of social contract – between two parties involved in an exchange. Now, it is well known that the Walrasian paradigm is void of exchange, and it cannot therefore handle this aspect of trading in rights. The conceptual framework of the Walrasian model is that, when the auctioneer has found his beloved equilibrium price vector, all the traders walk out into the market place, dump their excess supplies in a heap marked j for the $j^{th}$ commodity, etc., for all subscripts over goods, and removes his excess demands from heaps marked with suitable subscripts. That is not an exchange between two contracting parties, nor is it really an organized market, where predetermined sellers arrive to trade with predetermined consumers.

The second reason for the inability of the typical Walrasian system to handle institutions as endogenous variables I believe lies in the non-convex nature of trading in property rights – decision-making rights are by their very nature discrete, and cannot be chopped up into little continuous pieces. Hence, the necessary continuity features of the system will not be there, and any existing equilibrium will therefore not be unique, for this reason alone, as well as for others specified below.

The third reason lies in the economic function of institutions, as I have specified it here. If the purpose of institutions is to change decision-making and property rights in such a manner as to change the incentives for displaying socially unwanted behavior, there must be such behavior in the first place. In the standard version of the Walrasian system, however, no such deviant behavior can ever occur, it is simply assumed away. I believe that this is the ultimate reason for the absence of institutions in the standard analysis – the problems of human behavior that economic institutions ultimately deal with simply do not exist in the Walrasian world.

We might, of course, inquire whether it is possible to amend the standard assumptions in a Walrasian setting to allow for differences in utility functions to obtain a distribution of behavior, including non-cooperative behavior. Note that if we were to formally analyze the implications of individual-specific transaction costs we would then have to index transaction costs over individuals, rather than over commodities and

markets, as is the case with the more common definitions of transaction costs. It would then follow that the resulting trades, including those that establish institutions through the exchange of property rights, would depend not only on the distribution of initial endowments, technology, tastes over commodities, but also on the distribution of individual-specific transaction costs along the cooperation spectrum, or on the actual individuals that happen to coexist in a particular society. This, I believe, is nothing but a formal way of stating a proposition that historians have always repeated and never have succeeded in making economists take seriously. The point is simply that if the distribution of cooperativeness is non-constant, but a dynamic element that depends on the individuals that happen to coexist in a particular society at a particular point in time, then we shall never be able to understand the particular solutions to particular problems arrived at in that historical context without specifying all the historical conditions, including the role played by particular individuals on the stage of history. To wit, suppose we could conceive of two societies identical in the following respects: the same technology of production, exactly the same resource endowments, the same individual tastes over commodities, and with the necessary continuity features associated with these conditions fulfilled. In a standard Walrasian system, this would then lead to the two societies having the same equilibrium price vector and the same allocation of resources – i.e., they would be identical societies. However, this would not be true if we also allow for differences in individual behavior, resulting in different willingness to cooperate. The result would be that two societies would have different transaction cost structures, and, hence, they would also have different institutions to deal with the costs arising from differences in individual behavior. If there is a turnover in the pool of individuals in society, we could then explain observed institutions only by a detailed historical study of the conditions ruling at the time and place – i.e., by doing the kind of work that historians traditionally excel in.

Stating the issue in the context of a Walrasian setting also reminds us of another possibility, that of multiple solutions. Societies have different enforcement mechanisms available for dealing with problems of controlling individual behavior, i.e., they have a catalogue from which they can make a choice of the tool best designed to deal with the immediate issue at hand. Suppose a so-

ciety wishes to limit property rights by declaring theft, murder, prostitution, and the use of narcotic drugs undesirable activities. To enforce such institutional rules, it may choose to rely on the police force of the state combined with punishments meted out by the court system. Alternatively, it is conceivable that it could rely exclusively on family enforcement, for example, by using the taxing power of the state to make any family whose members either steal, murder, prostitute themselves, or use drugs pay an amount of taxes equivalent to the perceived damage done by the breaking of the rule. Or it might use various mechanisms for letting peer groups detect, report, and enforce the rules – ideas that are quite prominent in a few modern societies today. Yet another alternative is the use of social ostracism – anyone who violates the rules is cast aside and set adrift. Or the society might enforce a strong religion that effectively makes everyone believe that any breaking of existing rules will lead to automatic punishment either for future generations or for the individual himself after his death. So even if the desired result is a particular change in individual decision-making rights for the purposes of making society function smoother and getting more out if its limited resources, there may be many alternative ways of achieving that desired result. They all represent different institutional arrangements, for the rights to make decisions over others and over oneself are different in the examples I just quoted. It is not irrelevant whether we let the responsibility for enforcement lie with the police and the courts, the family, peer groups, social opinion, or the priesthood of our established religion. Whichever alternative or combination of alternatives a society chooses will imply very different institutional settings, even if the results in terms of eliminating undesirable behavior is attained equally efficiently.

Yet another reason for believing that there is no such thing as a unique set of efficient institutions lies in the concept of property rights itself. With this concept we usually understand the right to use as asset, the right to exclude others from laying claim to it, and the right to alienate it in various ways, such as by gift, bequeath, sale, or exchange. With private property we mean the right for a particular individual to use the asset, his individual right to exclude others, and his individual right to disposal. With collective property we mean the right of everyone to use, the right of no-one to exclude, and the right of no-one to alienate the asset. There are intermediate forms,

such as collective-exclusive, where a group retains the right of the members to use, the right of each member to sell his several right to usership, and the right of the group to exclude others. Or there is the notion of fee entail where an individual has the right to use it as he personally sees fit, the right to exclude others, but not the right to alienate the asset.

I stress these rather obvious points only to underscore the very important function of property rights to serve as the social tool for effecting different income distributions – in addition to the incentive mechanisms so often noted in the property rights literature. The different kinds of property rights just enumerated are different in both these aspects – in the incentive structures they afford for efficiency, and in the distribution of income that is implied by each of them. It therefore follows that, if we mean by institutions different property rights assignments, as I have suggested, two otherwise identical societies – i.e., identical with respect to tastes, endowments, and technology – with different individual behavior will differ not only in the institutional structure, but also in the resulting income distribution. If the society relies on police powers and court systems, lawyers, police officers, judges, and prison attendants will find increased demand for their services; if it relies on a strong religion, the priesthood will find itself in command of perhaps vast economic resources. Hence, institutions and income distribution issues are inseparably tied together, as an immediate consequence of the tie-in between property rights and institutions.

The implication of this relationship between the nature of property rights and the function of institutions is that the question of whether there exists a set of efficient institutions logically implies the question of whether there exists an optimal personal income distribution. Traditionally, economics has shunned this question, on the argument that personal income distribution is a question for the political, i.e., non-economic, part of the system to solve. If we accept the notion of an economic institution as a particular set of attenuated property rights, then this separation is no longer feasible. Efficiency and distribution can no longer be treated as different issues if institutions are endogenous variables, for institutions inevitably affect both. If we accept, as we have strong reason to, that there can be no optimal income distribution separate from ethical and moral considerations, i.e., normative value judgments then I think we shall also have to accept the pro-

position that the question of what institutions are efficient turns out to be a normative question – it depends on what income distribution you wish to argue for when you decide what institutions you wish to implement in a society.

I believe that there are some forceful implication stemming directly from this simple line of reasoning. Modern economic welfare theory relies fundamentally on the principle of separation of income distribution questions and issues of resource allocation. The proof of the optimality of the Pigouvian tax rules as methods for dealing with externalities, for example, relies implicitly on the proposition that a political authority can attain any income distribution it desires, and that in using the Pareto principle for judging policies towards externalities we can effectively disregard income distribution problems completely. The case is exactly the same with monopoly and public goods – here economists feel free to propose active policies without relying on any stronger value judgements than those inherent in the acceptance of the Pareto principle, and again, they feel free doing so on the basis of the idea of lump sum redistributions that can effect any politically desirable income distribution. However, if the connection between institutions and income distribution is as intimate as I have contended in this paper, then this separation theorem has a more limited applicability than is generally recognized.

First, we have to make a basic distinction between two policy tools of the government – altering incentives by changing relative prices, and changing incentives by altering institutions and property rights. Insofar as lump sum income redistributions are ever possible, which I am prepared to accept for the sake of the argument, I believe that such policies as affect only relative prices do conform to the separation principle. Quite possibly, Pigovian taxes as a means of dealing with externalities fall into this category. If we tax a paper and pulp mill at Silver Bay in Minnesota for polluting Lake Superior, we may effectively attain a reduction in the outpour of mercury and other pollutants at the cost of severe unemployment; however, by appropriate income maintenance schemes we may compensate the losers from this, even the firm owners should we so wish, and effectively attain the same income distribution as before. However, this does not appear to be the case when it comes to changing institutions, i.e., property rights; furthermore, I believe that there is a basic assymetry in the implementation of institutional changes for altering indi-

vidual behavior. To exemplify, suppose we wish to preserve the redwoods of northern California, but by prohibiting logging of the trees we take away the livelihood from the people in Eureka – a redwood logging town. These people may be compensated, in spite of this change in the institutional setting. That is to say, we can compensate for the loss of income due to prohibitions, and so the separation principle probably holds. On the other hand, I do not think the separation principle holds for changes in property rights that create new rights. As an example, take the opening up of federal land for the mining of coal. This is a change in property rights, and an institutional change. We now allow mining where it used to be prohibited. In so doing we create the right to exploit a mineral resource and to derive income therefrom. This is not a change in relative prices, it is an income redistribution from those who would have grazed on the land or used it for recreation, who may be compensated should we so wish, to those who derive the income from coal mining. However, should we consider the income to the miners excessive, there is little that can be done about it. In order to exploit the resource, income incentives, i.e., property rights allocations, have to be used, and there is then no possible method for restoring the previously existing income distribution. We cannot tax the miners without simultaneously reducing the incentives for mining, and so income distribution and allocation of resources cannot be effectively separated. The assymetry thus seems to be that we can make such a separation for the destruction of rights, but not for the creation of new rights. Hence there are instances where the fundamental theorem of modern welfare economics does not seem applicable.

If what I have said so far seems more destructive than constructive, in the sense that I have stressed a number of observations that would mitigate against a belief that we shall ever be able to build purely economic theory of institutions, I believe that it is now time to temper this message somewhat. If institutions arise through voluntary agreements in a market-like context, as I have suggested as a perhaps fruitful point of view, then we ought to see similar solutions to similar problems no matter what the overall institutional background in terms of political, religious, and cultural enforcement mechanisms a society may choose to implement. I have repeatedly referred to the firm as such an example of an institution that

arises spontaneously through the interaction of self-interested parties in a market situation; there is thus little need to belabor that obvious point.

However, I merely wish to add to this the observation that the relationship that we have come to call the firm is not limited simply to capitalistic economies. Even if we accept the argument by Alchian and Demsetz that the kind of transaction costs that the firm serves to minimize are those associated with the organization and monitoring of team production, we are still not necessarily bound by their conclusion that they have explained only the workings of the typical capitalistic unit of production. I believe their argument to be much more general than that, for the problem of making an individual worker behave according to implied or explicit clauses in a labor contract arises, and seems at least equally prevalent, under collective ownership and organization of production, such as in contemporary socialist economies. Thus, organizations for the monitoring of team production, very similar in their function to the capitalist firm, will arise even in planned economies, and the only major difference would appear to lie in the manner in which the monitor receives his compensation – he is scarcely a residual claimant in planned economies with state ownership of productive resources, unless, of course, we wish to consider the state the residual claimant, in which case the parallel becomes complete. The simple point I wish to make is, however, only that there are problems which seem to command a universal solution, no matter what the superstructure in the form of political mechanism and resource ownership may be. Thus, whatever other differences they may have, societies like the capitalistic western economies and the industrialized planned economies face similar problems in the monitoring of team production, simply because they use similar resource endowments, have similar processes, and similar human beings – and so the institutional solutions will be similar, since the transaction costs problems they must deal with in order to organize production efficiently will also be similar. Thus, whereas all institutions ought to be treated as endogenous features of an economic system, some institutions appear as more endogenous than others. Or perhaps it is better to say that the instances in which there are multiple solutions or equilibria are not universal, i.e., there really are at least partial solutions that are unique.

## IV. Conclusions

I have offered some glimpses of a method of thought rather than a finished model. I have tried to show how the property rights – transaction costs approach can serve as a rich and potentially important addition to standard economic theory by providing us, at least, with an appropriate language for the analysis of economic institutions. From the choice of this language and its implied definitions will follow some immediate implications, and I have tried to stress some of them in the preceding. Of these, I would think the lack of uniqueness of institutions and the role of institutions as a means of attaining income redistribution are perhaps the most important. I would not for a moment suggest that my previous discussion in any way is exhaustive – there is more water in this well, and it may be that I have only found the muddy parts of it, so the clear and fresh may still be available for future use.

## Footnotes

* University of Wisconsin, Madison.

1. A A Alchian, *Pricing and Society,* Institute for Economic Research, occasional paper # 10, 1967.

2. R H Coase, "The Nature of the Firm", *Economica,* 4 (1937) 386–405.

3. A A Alchian and H Demsetz, "Production, Information Costs, and Economic Organization", *American Economic Review,* 62 (1972) 777–7795.

4. C J Dahlman, "The Problem of Externality", *Journal of Law and Economics,* 22 (1979) 141–162.

5. C J Dahlman, *op cit.*

6. C J Dahlman, *The Open Field System and Beyond,* forthcoming from Cambridge University Press.

7. C J Dahlman, "Transaction Costs, Money, and Bonds", unpublished working paper.

8. O Wiliamson, *Market and Hierarchies,* Free Press, New York, 1975.

9. B Klein, R G Crawford and A A Alchian, "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process", *Journal of Law and Economics,* 21 (1978).

10. H Demsetz, "Towards a Theory of Property Rights", *American Economic Review,* 57 (1967) 347–350).

11. C J Dahlman, "Transaction Costs, Money, and Bonds", *op cit.*

12. See my "Collective Ownership and Economic Organization" for a discussion of this.

# Population Pressure and Agrarian Property Rights in Haiti

BY MATS LUNDAHL*

*This paper demonstrates that population pressure on the land has been an important determinant of agrarian property rights in Haiti, including property rights in human beings. Major changes in the density of the population are identified and linked to redefinitions of property rights. The paper ends with a discussion of possible future monopolization of landholdings in Haiti.*

## Introduction

Furubotn and Pejovich define *property rights* as "the sanctioned behavioral relations among men that arise from the existence of things and pertain to their use," and argue that the "prevailing system of property rights in the community can be described, then, as the set of economic and social relations defining the position of each individual with respect to the utilization of scarce resources."[1] For a long time, property rights constituted a neglected field in economic theory – presumably, mainly as a result of the increasing mathematization of the discipline after the Second World War. During the past decade, however, the concept has played an increasingly important role in economic research, especially in attempts to link theory with empirical evidence. According to Furubotn and Pejovich, the aim of the property rights approach to economics is to establish operationally meaningful, i.e., empirically testable, propositions about the economy, given postulates on maximizing behavior and the sovereignty of individuals' preferences or values in guiding economic choice. For such an approach to yield fruitful insights, the institutional environment within which economic activity takes place must be specified with great care.[2]

The development of property rights and institutions can itself be subjected to economic analysis. This paper attempts to link the concept of property rights with the degree of population pressure on the land in the setting of an underdeveloped agrarian economy: that of Haiti. It will be shown how changes in population pressure, and, hence, in relative factor supplies, have constituted an impor-

tant determinant of the system of property rights in Haitian agriculture from the French colonial period up to the present time. In this context, the "things" referred to by Furubotn and Pejovich are not only land but also men – those men who work the land. The major changes in population density have been linked to important changes not only in relations between the laboring and non-laboring classes, which pertain to the use of agricultural land, but also in the relations connected with the use of labor. Property rights in both land and human beings have been redefined and the degree of population pressure has had an important role to play in this process. This has not been a role which could be unequivocally predicted from a known man/land ratio or from a change in this ratio, but one which has differed as a number of circumstances exogenous to population growth have differed.

## The Rise of the Plantation System

Present-day Haiti was a French colony from 1697 to 1791. With respect to property rights, the main characteristic of Saint-Domingue, as the colony was known, was the combination of large-scale plantations with slave labor. By the time of the French Revolution, some 450,000 Negro slaves[3] were sustaining an economy which produced a number of export crops, notably sugar and coffee, on fairly large-sized plantations. The largest were the sugar plantations which ranged from 150 to 300 hectares, while coffee and indigo plantations were usually less than a hundred hectares.[4] Sugar was the most important crop. It was basically the technical requirements of sugar production in

combination with an exteme demographic situation which produced the system of property rights that prevailed in Saint-Domingue and then, in a slightly modified form, in independent Haiti for more than a decade after liberation from the French.

Sugar cane came to Hispaniola at the beginning of the sixteenth century. The West Indian climate presented extremely favorable conditions for its cultivation and Hispaniola and the rest of the Caribbean islands possessed a strong comparative advantage in sugar production. The structure of this advantage was not such, however, that it could be acted upon directly. The technology available to the sugar planters required comparatively heavy concentrations of capital, land and labor for profitable operations.[5] Each plantation had to have a crushing mill, the optimum economic size of which was fairly large. This optimum, in turn, determined the optimum size of the plantation and the size of the required labor force.

Establishing plantations of the requisite size was easy, but the recruitment of the necessary labor force presented a formidable problem for the planters. The reason is to be found in demographic changes. When Columbus discovered Hispaniola in 1492 the island sustained a large Indian population, estimates of which range from 200,000 to 1,200,000.[6] A century later hardly a soul of this population was left. Spanish practices of forced (encomienda) labor in combination with imported European diseases and outright slaughter in battle had taken a heavy toll. This meant that there was plenty of land to turn into plantations, but also that Negro slaves had to be imported from America-to man them. This practice was already underway in 1502, but it was not until the French period that the slave traffic reached its peak, with average annual imports possibly exceeding 20,000 people.[7]

The dwindling population was less of a problem during the Spanish period, when extraction of alluvial gold was the Spaniards' principal economic interest. When this activity ceased, cattle grazing and livestock trade became the dominant activities occupying this position as early as the 1530s and 1540s. Cattle ranching is a highly land-intensive acitivity requiring very little labor with the cattle being allowed to stray across vast open ranges. The basic economic units in this system were the hatos, "immense possessions. . . where horses and cattle [were] raised with little care."[8] By 1650, hatos may have covered as much as one-third of the area of Hispaniola.[9] Cattle

ranching was quite in harmony with the factor proportions prevailing in the island, but after the formal cession of Saint-Domingue to France by the Treaty of Ryswick in 1697, this equilibrium was upset. The French colonists, who had penetrated western Hispaniola at least seventy-five years earlier, had been hesitant to undertake the large investments required to make sugar cane a profitable crop as long as the territorial status of Saint-Domingue remained uncertain. With the Spanish threat removed, sugar cultivation was expanded rapidly across the French colony. Large plantations emerged and property rights were created not only in land, but also in men as an artificial means of overcoming the obstacle posed by an extreme demographic situation where plantation labor on a voluntary or forced basis was unavailable locally.

It must also be mentioned that, although the slavery-based plantation system was the "final" solution of the labor force problem, it was not the only one attempted. At an earlier stage, indentured laborers (engagés) had been brought in from France on contracts specifying the number of years (usually three) they had to work before gaining complete freedom.[10] This, however, was no solution to the problem of mobilizing labor for the sugar estates. The coercive measures at the disposal of the planters vis-à-vis the engagés were too weak. An indentured laborer could be held only for a limited number of years, could not be driven as relentlessly as a slave and when the labor contract expired he was a free man. With plenty of unsettled land available, backbreaking labor on a sugar estate would never have attracted a single ex-engagé. The necessary effort could only be extracted from slave labor. For sugar to be a profitable crop, extremely strong property rights had to be established not primarily in land, which was plentiful, but in human beings.[11]

## A Nation of Free Peasants

After the French were expelled from Haiti, slavery was abolished in 1793 and fifty years later Haiti was a nation where free peasants were making an independent living on land which belonged mainly to them. During the intervening years, Haiti's entire economic system had been profoundly reshaped. The set of property rights had undergone a fundamental change and once again one of the main determinants of change was to be found in the relative availability of production factors.

Based as they were on slavery, the large plantations – *la grande culture* – gradually disappeared after 1793. The institution of slavery could be upheld only as long as there was a supply of slaves and slaves were available only as long as effective sanctions excluded the Negro masses from other ways of making a living. After independence, the range of opportunities increased as a result of the drastically increased availability of land together with weak public administrations.

The transition from slavery-based plantations to free peasant smallholdings did not take place immediately after the end of French rule. On the contrary, the first rulers of independent Haiti felt that the plantation system should be preserved and made strenuous efforts to retain it.[12] Much of the system had been physically destroyed during more than ten years of intermittent warfare, but enough was left for a restoration of *la grande culture* to be a feasible option. Thus, up to the historically significant year 1809, the year of the first land reform in Latin America, an agrarian system which differed from slavery in name only existed in Haiti. Ex-slaves who during the turmoil produced by the wars of liberation had been acting mainly as independent agricultural small-scale producers were, as far as possible, brought back to the estates. The plantations were rented to members of the emerging Haitian elite and strict military supervision of the agricultural workers was resorted to in order to secure the necessary labor input.

The restoration was only a temporary episode, however. In 1809, Alexandre Pétion, president in the southern half of the country, decided to set his serfs free and to redistribute the large landholdings. Ten years later, Henry Christophe followed suit in his northern kingdom. By 1840 Haiti had become a nation of free peasants and this situation was to be reinforced during the rest of the nineteenth century. In thirty years, the system of agrarian property rights had been completely transformed. No one now held any rights in his fellow men, and, one way or another, the peasant population had access to land which they could till for their own benefit: as outright owners, as squatters or as sharecroppers. This constituted one of the most decisive events in Haiti's economic history. The creation of an economy comprised of free peasants set Haiti on a course which diverged widely from the pattern typical of most of Latin America.[13]

The transition from plantations to peasant holdings can be traced to a large extent to the chang-ing effective supply of labor and land.[14] To understand how this worked we may take a brief look at the phenomenon of *marronage*. During the colonial period this term referred to the escape, organized or unorganized, of slaves from the French plantations. These runaway slaves fled to remote regions outside the effective control of the colonial administration where they attempted to make a living as subsistence farmers.[15] *Marronage* never developed into a mass movement. Its success was ultimately conditioned by the amount of land available for illegal squatting without interference by the authorities and during the colonial period this area was limited in practice. The planters and their administrative machinery were sufficiently strong to ensure that *marronage* was a solution for a minority of dissatisfied slaves only. Policing expeditions were regularly sent out when it was felt that the strength of the maroon communities exceeded the tolerable level.

During and after the wars of liberation the extent of *marronage* increased.[16] When the French administrative apparatus had been destroyed and the balance of power no longer weighed so heavily against the Negro masses and when, in addition, many colonial plantations had been abandoned and lay without effective ownership, the area available to those ex-slaves who preferred independent subsistence farming to militarily supervised serfdom increased. Now, the masses were provided with an attractive alternative to remaining as landless workers on plantation estates with a rigid discipline.

The increased availability of land had important repercussions in the labor market. During the wars most ex-slaves, when given a choice, preferred to work on their own small plots instead of going back to the plantation system. In Saint-Domingue part of the slaves' subsistence was secured by providing them with small garden plots, the produce of which the slaves could dispose of themselves, in markets or by direct consumption. To a certain extent, these "provision" plots provided the colonists with foodstuffs.[17] During the wars of liberation when imports of food or their distribution within Saint-Domingue were disrupted, the food supply gradually came to depend on the "slave gardens", and it appears that a very widespread reaction among the ex-slaves was simply to remain as cultivators on their "old" plots.[18] Presumably, it was the very knowledge of this which made the first Haitian rulers take the decision to reinstitute the plantation system on a *forced* labor basis. Any attempted solution based on a

free choice would have failed. On the macro-econ-omic level, the situation was turned further a-gainst the plantation system by the population de-cline which eventually resulted from the wars. From 1790 to 1805 the Haitian population de-clined by an estimated 150,000.[19] Of these, 40,000 were whites [20] but the majority of the remainder were Negro ex-slaves.

Thus, the relative supply of land had increased, while that of labor had decreased. Only artificial administrative devices could for a time guarantee the survival of the "colonial" pattern of property rights and when the administrative apparatus had been sufficiently weakened the pattern broke down. Toussaint and Dessalines could muster enough strength to keep the old military system working for some time. Since no agreement had been reached with France regarding the territorial status of the country, the threat of renewed war activities had not been removed when Dessalines was murdered in 1806. When Haiti was divided into two states intermittently waging a civil war upon each other following his death, successively less money and energy could be spent on pre-venting the system of property rights from falling apart. At the same time the costs of supervision and enforcement had increased. The masses had tasted freedom during the revolutionary wars and were less prepared than ever to go back to the plantations. The beginning of the end came in 1809 in the southern part and ten years later in the north. Although there were occasional at-tempts,[21] no subsequent Haitian administration was able to reverse this order of things. The new system had come to stay. When rights were se-riously threatened, rural protest movements arose which sometimes turned into outright peasant re-volts.[22] Subsequent changes in the agrarian prop-erty rights system have been modifications within the peasant mode of production rather than pro-found transformations involving the relative freedom of men.

## Securing Peasant Ownership

By 1842 probably none of the colonial plantations remained in their original form. Around one-third of the population were peasant-owners, another third were squatters and most of the remainder were sharecroppers.[23] All of them were smallhold-ers. This distribution of land did not remain un-changed, however, during the nineteenth century. Before 1900 the majority of the Haitian peasants could probably safely be termed "owners". The

reason for this development was the compara-tively strong bargaining position conferred on the peasants by the low man/land ratio. In 1978 the population density amounted to 174 persons per square kilometer.[24] In the 1820s, the maximum figure was 25,[25] a figure which was to increase only slowly during the course of the nineteenth century. To see how peasant "owners" came to dominate the scene we will outline more details of the change from large plantations to smallhold-ings.

When the attempt to preserve the colonial plantations was made at the beginning of the nine-teenth century most of the land was declared government property and, thereafter, was rented to high army officers and other members of the new upper class. As the colonial property rights structure finally began to crumble under Pétion and his successors this government property was transferred to private hands. At the same time the landed elite found that plantation labor was no longer available in the quantities and on the conditions necessary for profitable operation and took steps to adapt to the changing circumstances. Since cultivating the soil themselves was out of question, the first option was to lease the land to the peasants against collection of some type of rent, usually a share-rent: one-half of the crop. This strategy was obviously feasible only for a limited period, however, because the interests of landlord and peasant often clashed when it came to the exact determination of contractual obliga-tions.

The area of conflict was in the physical har-vesting of the crops planted under sharecropping arrangements.[26] In contemporary Haiti the sha-ring arrangement means that a division of the rented plot is made before the plot is harvested and the landlord himself must harvest his half and see that the produce is marketed. Presumably, the same type of arrangement became the rule in nine-teenth-century Haiti.[27] This posed a very obvious problem for the landlord class:

Except for the owners of coffee plantations...,[28] the nineteenth century Haitian landlord was in the almost ludicrous position of having fields cultivated in crops which did not really interest him, and of having fur-thermore to harvest those peasant crops himself. As a last straw he was also obliged, if he was to make any money of the arrangement, to himself arrange for the marketing of that produce within the arena of a popular market system dominated by energetic female peasants. If the image of a self-respecting member of the gentry

digging up his own sweet potatoes is humorous, the image of his genteel, French-speaking wife lugging them to a local market to sell them in noisy competition with skillful peasant *machând* is absurd.[29]

Thus, except for the case of coffee, this type of sharecropping contract was more or less doomed from the beginning. Only during the first few years after the initial redistribution of land would we expect to meet it and then presumably with the sharecropper harvesting the portion of the landlord as well. Unfortunately, no statistics have so far been uncovered to support the hypothesis, but it seems reasonable to expect that a majority of the sharecropping contracts in 1842, referred to above, dealt with coffee plantations.

Due to the comparatively easy availability of land for cultivation the Haitian peasants were in a much better position to oppose landlord claims than their counterparts in most parts of the world. Sharecropping was not a viable solution from the point of view of the landlords and, therefore, it gradually disappeared. Instead, the predominant pattern became one where the peasants actually *owned* their fields – generally without deeds. This situation arose in two different ways, by laissez-faire squatting and by alienation of parcels by the landlords through actual sales.

The bulk of the literature on Haiti puts the emphasis on the importance of squatting.[30] When the landlords found that going back to the plantation system was impossible and that sharecropping was not viable, they simply gave up, withdrawing to an urban life and allowing their tenants or other peasants free reign. Recently, however, Gerald Murray has strongly challenged this traditional view and pointed to the possibility that most peasants actually acquired their land via regular purchases based, on the one hand, on the need of the landowning group to capitalize on land which its members did not want to cultivate themselves and for which no hired fieldhands could be found, and, on the other, on cash accumulated by the peasants from transaction in the domestic marketing circuit.[31]

Murray's interpretation_is_interesting since it simultaneously provides an explanation of why in spite of a general absence of written titles, peasant holdings appear to have been fairly secure and highly marketable in Haiti.[32] A sales transaction should constitute a firmer basis for both tenure and further transactions than simple squatting. Two more considerations could, however, be added here. In the first place, the sales to which

Murray refers took place during a period when land was plentiful in relation to the population. There was enough land for everyone who wanted a plot at least up to the last quarter of the nineteenth century.[33] In this situation, few people, (and especially not outside interests), were likely to question even unwritten land rights, since the labor necessary to produce an income from the land was lacking.

The second point is one which has relevance also to the contemporary situation. Rural Haiti is a relatively classless society[34] and a vast majority of all land transactions take place within the context of the rural world, i.e. between people of basically the same social standing – people sharing the same values. Such people are not likely to question the rules of a game which has evolved within more or less the same setting during a century and a half. The situation would be different if rural Haiti had been socially highly stratified and land transactions had been carried out mainly on an *inter*class rather than an *intra*class basis.

By the end of the nineteenth century the transition from slavery-based plantation to a society where the vast majority of cultivators were peasants who owned the land themselves had been completed. The history of the development of property rights during the twentieth century is not well known. No cadastral survey has even been undertaken in Haiti which can shed light on the contemporary situation. An attempt was made during the America occupation of the country (1915–34) to straighten out the land tenure situation, presumably to prepare the way for American-owned plantations. Aerial photography was carried out but before the photographs had been interpreted the building where the negatives were stored burned down "in an unexplained fire".[35] To evaluate today's situation, we are left with the rather unreliable figures of the 1950 census and a number of local surveys.

Presumably, however, no major changes have taken place. The available information is difficult to interpret but, in the main, it indicates that a majority of all Haitian peasants own their land. The 1950 census indicated that up to 85 percent of the peasants were "owners". This impression is confirmed by at least two later major surveys, one nationwide in 1970 and another of more than 7,000 farmers in the *arrondissement* of Cap-Haïtien in 1974, in which it was found that 60 percent of all parcels and 75 percent of all the land, respectively, were cultivated by the owners themselves.[36] According to all three sources, the in-

cidence of tenant farming and sharecropping was low: some 8 percent in 1950 (peasants owning *no* land – not part-time tenants), 28 percent of all parcels in 1970, and 14 percent of the area in the 1974 Cap survey. The generally accepted picture of today's landholding system in Haiti is that a majority of the peasants still own their land with or without deeds and that most of the land *area* is held in this way.

## Future Monopolization of Land?

Let us end with a brief look at the future. From the mid-nineteenth century up to the present time Haiti has stood out as an exception to the land tenure pattern prevailing in most Latin American states. Land has not been concentrated in the hands of a minority while the mass of the rural population have been landless laborers, tenants or minifundistas working on artificially overcrowded marginal soils. Haiti has not had any "land problem" in that sense. Rather, the main difficulty has been to maintain fertility on fairly equitably distributed plots in the face of population growth. In this struggle, the Haitian peasant has generally not been successful,[37] but at least, one can claim, he has been spared exploitation by a landlord class. Can we expect the same pattern to continue into the future or will the growth of the Haitian rural population lead to dramatic shifts in the structure of agrarian property rights towards increasing concentration of land and, hence, also to exploitation of the landless?

Most of the literature which deals with the possibility or existence of land concentration in Haiti is concerned with attacks on peasant freedom by a class of absentee landlords. A number of authors have, in fact, attempted to prove that such a concentration of land already exists in Haiti.[38] However, such an interpretation violates the observable facts.[39] Haiti *de facto* is a country where most of the rural population has access to land on terms which cannot be qualified as monopolistic.

What then is the likelihood of the emergence of such a class? It is well known that very few Haitian peasants can present any written titles to their land.[40] Furthermore, Haitian history points to a number of instances where, when the value of the land has increased, peasants have been subject to eviction by outsiders.[41] However, such cases must be considered rare. Murray found that in a community he studied in depth this had *never* occurred.[42] The main reason appeared to be that although very few peasants could present indi-

vidual titles to the plots they owned, the *grâ-pyês* generally still existed. This "big" deed to the undivided land of a family estate some generations ago was kept by some relative and could be used to trace subsequent land transactions.[43] The existence of such documents undoubtedly makes alienation of peasant land difficult for outsiders.

A problem may arise even among those possessing legal deeds. Unwritten property rights, as we have already pointed out, are generally regarded as valid by the peasant class from which potential "insiders" would come. This convention is reinforced by a second factor based on sorcery. In Murray's community the most important threat to peasant security was not seen as coming from outsiders but rather from distant kin who actually *did* have legal rights to land but who, by emigrating or otherwise, had in practice forfeited their rights. In such cases it is, of course, possible that generations later heirs could come up with a legal title. This type of intruder was, however, regarded as being particularly vulnerable to sorcery exercised by those actually *using* the land.[44] Thus, the likelihood that people with legal rights to land which they had chosen to leave would come back to claim that land seems low.[45]

Population growth may possibly disrupt this relative security. One such pattern has been suggested by Murray. His point of departure is that an individual who can today buy land in Haiti will never lack the labor to make the land productive and, hence, to make the transaction worthwhile.[46] This, according to Murray, is ensured by the existence of potential sharecroppers. He then goes on to argue that

it is precisely such a situation which is conducive to the emergence of patterns of land concentration. Such a danger would exist no matter what the pre-existing tenure mode were. But in a society such as Haiti, where even at a "grass roots" level land has traditionally been alienable, the danger is especially great. For where there is land purchase, there must also be land sale and – ipso facto – the emergence of at least temporary resource differentials. And where land is further transmitted via inheritance, as is true of Haiti, these differentials will easily be intergenerationally perpetuated. Furthermore, since the children of the better-off start life in a somewhat stronger economic position than the children of the less well off, they are more likely to purchase more land, the differentials will thus increase, and land concentration will have set in.[47]

This has not occurred so far, however, because there is another mechanism which serves as a pe-

riodic regulator of the distribution of land, namely, voodoo. For reasons connected with the need to finance voodoo ceremonies at various times over the life cycle, land has to be put on the market for sale. Murray found that a majority of all land sales in the community were motivated by these needs. The result of these transactions, as Murray sees it, has been to reduce class differentials based on land tenure:

The mechanism has not eliminated differentials, but it has kept them within the basic confines of a life-cycle modality of resource management, and has prevented the emergence of intergenerationally perpetuated local strata.[48]

Such a view of the land market is highly dubious. This double role as a generator and moderator of class differences is definitely not inherent in the market mechanism. *A priori* there is no reason to expect that those with more land are better farmers who will improve their economic positions and, therefore, also buy more land.[49] Neither should we expect that those with less land are necessarily the main sellers. Finally, there is nothing in the market mechanism which guarantees that the land coming from voodoo-induced transactions is land which is alienated by those holding relatively much land. All these propositions have to be proved before Murray's case can be established.

Perhaps the most realistic type of mechanism based on population growth which may eventually undermine the prevailing set of agrarian property rights in Haiti is to be found in increasing poverty itself.[50] There is a tendency for rural incomes to fall over time. So far, one of the main regulators here has been migration to the capital city and abroad. In the future there are, however, no guarantees that emigration to other countries will continue to provide a safety valve. It may very well be that other countries feel that too many Haitians are coming in and they may, therefore, take steps to curtail immigration.[51] In such a situation greater stress will be placed on the domestic economy to provide the population with non-agricultural employment. Hitherto, the economy has failed to do so. If the rural population continues to grow, marginal peasants may find themselves in a situation where they have to increase their indebtedness with land as collateral and this may lead to an eventual transfer of land into the hands of moneylenders. Alternatively, land may have

to be sold to cover immediate needs. This is a familiar pattern in other agrarian communities.[52]

Concentration of land tends to lead to monopsonization of the labor market. When large segments of the population lack land of their own they become increasingly dependent on landowners for employment. In this situation exploitation may be a reality.[53] Whether such a situation will develop in Haiti remains to be seen. So far, nothing indicates that it is imminent, but it may be prudent to concentrate some attention on uncovering possible hidden or unknown trends in the development of agrarian property rights. One cannot simply trust the market mechanism since there is, of course, nothing inherent in that mechanism which guarantees that the development of property rights takes the most "desirable" course. In this sense the market is neutral. It all depends on the circumstances under which the market mechanism is allowed to work.

## Footnotes

* University of Lund. Thanks are due to Carl-Johan Dahlman, Lennart Jörberg, Bo Larsson and Jim Love for their constructive criticisms of an earlier version of this paper.
1 Furubotn and Pejovich (1972), p 1139.
2 Ibid, p 1157.
3 Moreau de Saint-Méry (1958), p 28.
4 Lepkowski (1968), pp 48–49.
5 For details regarding the sugar economy see Lundahl (1979), pp 256–59.
6 The estimates of the indigenous population vary widely from source to source. For a sample see e.g. Palmer (1976), p 38, Cauvin (1977), p 39, Lundahl (1979), p 189, Caprio (1979), p 28 and the sources indicated in these work. Cook and Borah (1971) discuss the aboriginal population of Hispaniola at length.
7 Lundahl (1979), p 189.
8 Moreau de Saint-Méry (1796), p 65.
9 Palmer (1976), p 51.
10 The *engagé* system is discussed in Debien (1952).
11 In this respect, the colonial economy is consistent with the Domar hypothesis regarding the causes of slavery or serfdom which states that out of free land, free peasants and non-working landowners, any pair of elements, but not all three, can exist simultaneously (Domar [1970]). For an efficient exploitation of the possibilities offered by sugar cane when land was plentiful and when landowners would not work themselves on the land, laborers had to be enslaved to prevent them from taking advantage of the easy availability of land. Also, ownership of land was monopolized by the free citizens of the colony.

[12] See Lundahl (1979), pp 259–63.
[13] This course is analyzed at length in ibid.
[14] See ibid, Chapter 6 for a discussion of all the factors involved.
[15] Extensive discussions of *marronage* can be found in Debbasch (1961), (1962), Debien (1966), and Fouchard (1972).
[16] Lepkowski (1968), note, p 80.
[17] Murray (1977), p 49.
[18] Ibid, pp 57–64.
[19] Lundahl (1979), p 272.
[20] Ibid, p 320.
[21] See ibid, pp 264–68.
[22] See Nicholls (1979), esp. pp 30–31.
[23] Leyburn (1966), p 76.
[24] Lundahl (1979), p 55.
[25] Franklin (1828), p 404.
[26] A second area of conflict, suggested by Murray (1977), pp 94–96, that of the choice of crop to be planted, is harder to accept. Murray argues that "the entire orientation of landowners of the period was the production of crops for export, above all the production of sugar cane which had underwritten so many colonial fortunes," (Ibid, p 94) while the peasants preferred to grow crops which could be sold via the internal marketing system with which the peasants were familiar since the colonial period. Exceptions here were coffee and cotton – "simply because the trees were already there. . ." (Ibid, p 95.) The sales of export crops were conducted via licenced government traders with whom peasant contacts were "disadvantageous and perhaps perilous," (Ibid.) while the marketing of domestic crops took place via a network of market women basically coming from the peasant class itself.

There are at least two difficulties with such an argument. In the first place, sugar quickly ceased to be an export crop in the post-independence period. With the technology of the period, as we have already discussed, sugar processing required high concentrations of capital, labor, and land, and such concentrations were simply beyond the means of the small peasant producers who rented the land. It is therefore not likely that the landlords would have insisted on sugar cane being grown, especially not since increased competition from Cuba and other Caribbean islands as well as from European beet-sugar made the price of sugar decline during the first half of the nineteenth century (Lundahl [1979], p 274). The second difficulty lies in the fact that an argument which holds that sales of export products are difficult due to the risks entailed in dealing with government licensed intermediaries and which simultaneously maintains that there was no conflict over the choice of crop in the case of coffee is self-contradictory, since coffee is the prime example of a crop marketed in this way. Rather, the absence of conflict in the case of coffee should have been due to the extremely low labor requirements connected with this crop.

The standard procedure was to leave virtually eve-rything except harvesting to nature. (Cf Lundahl [1979], pp 236–37, 564–65.) Hence, the attraction of coffee for the peasants was that it could be cultivated without much labor effort and still yield an income to be added to that resulting from the cultivation of foodstuffs. (Murray employs the latter argument as well but attempts to reconcile it with that of the choice of marketing channels.)
[27] Murray (1977), pp 96–97.
[28] Cf. note 26.
[29] Murray (1977), p 97.
[30] Especially the highly influential works by Leyburn (1966), pp 76–79, and Moral (1961), pp 27–28. Cf. also Lepkowski (1968), pp 120–21.
[31] Murray (1977), pp 107–08.
[32] Cf. ibid, pp 349–54.
[33] Ibid, p 410.
[34] For discussions of the Haitian class system, see the numerous references quoted in Lundahl (1979), note 83, p 361.
[35] Schmidt (1971), p 179.
[36] Lundahl (1979), pp 48.
[37] This is the main theme in Lundahl (1979). Cf., however, also Palmer (1976), pp 167–71, for an exception to this pattern.
[38] E.g. Casimir (1964), Brisson (1968), Pierre-Charles (1969), Jean (1974).
[39] Cf. Lundahl (1979), pp 51–52, Zuvekas (1978), pp 92–98.
[40] According to Murray (1977), p 351, probably fewer than one percent.
[41] Lundahl (1979), pp 603–04.
[42] Murray (1977), p 352.
[43] Ibid, pp 310–11, 352–53.
[44] Ibid, pp 320–22.
[45] Palmer (1976), p 149, however, reports the opposite pattern, where those remaining in the countryside do not dare touch fallow land owned by people who have left the community.
[46] Murray (1977), pp 463–65.
[47] Ibid, pp 463–64.
[48] Ibid, p 465.
[49] In economies of the Haitian type, there is frequently a low correlation between the initial wealth of a person and his entrepreneurial abilities. Cf. McKinnon (1973), p 11.
[50] Cf. Lundahl (1979), pp 645–46.
[51] Emigration from Haiti is dealt with in ibid, pp 623–28 and Zuvekas (1978), pp 73–76.
[52] Cf. Myrdal (1968), pp 1039–47.
[53] Cf. Griffin (1976).

## Bibliography

Brisson, Gérald. *Les relations agraires dans l'Haïti contemporaine.* Mimeo. México, D.F., 1968.

Caprio, Giovanni. Haiti – wirtschaftliche Entwicklung und periphere Gesellschaftsformation. Frankfurt/Main. 1979.

Casimir, Jean. "Aperçu sur la structure économique d'Haïti", *América Latina,* vol. 7, 1964.

Cauvin, Henri. "The Haitian Economy: A Case Study of Underdevelopment", PhD thesis, New School for Social Research, New York, 1977.

Cook, Sherborne F. and Borah, Woodrow. "The Aboriginal Population of Hispaniola" in *Essays in Population History.* Berkeley, 1971.

Debbasch, Yvan. "Le marronage: Essai sur la désertion de l'esclave antillais", *L'Année Sociologique,* vol. 3, 1961, 1962.

Debien, Gabriel. *Les engagés pour les Antilles (1634–1715).* Paris, 1952.

Debien, Gabriel. "Le marronage aux Antilles Françaises au XVIIIe siècle", *Caribbean Studies,* vol. 6, 1966.

Domar, Evsey D. "The Causes of Slavery or Serfdom: A Hypothesis", *Journal of Economic History,* vol. 30, 1970.

Fouchard, Jean. *Les marrons de la liberté.* Paris 1972.

Franklin, James. *The Present State of Hayti (Saint Domingo) with Remarks on its Agriculture, Commerce, Laws, Religion, Finances and Population, etc. etc.* London, 1828.

Furubotn, Eirik G. and Pejovich, Svetozar. "Property Rights and Economic Theory: A Survey of Recent Literature", *Journal of Economic Literature,* vol. 10, 1972.

Griffin, Keith. *Land Concentration and Rural Poverty.* London and Basingstoke, 1976.

Jean, Rodrigue. *Classes sociales et sous-développpement en Haiti.* Ville St-Laurent, 1974.

Lepkowski, Tadeusz. *Haití. Tomo I.* La Habana, 1968.

Leyburn, James G. *The Haitian People.* Revised edition, New Haven, 1966.

Lundahl, Mats. *Peasants and Poverty: A Study of Haiti.* London and New York, 1979.

McKinnon, Ronald I. *Money and Capital in Economic Development.* Washington, D.C., 1973.

Moral, Paul. *Le paysan haïtien (Etude sur la vie rurale en Haïti).* Paris, 1961.

Moreau de Saint-Méry, Médéric Louis Elie. *A Topographical and Political Description of the Spanish Part of Saint Domingue.* 2 volumes, Philadelphia, 1796.

Moreau de Saint-Méry, Médéric Louis Elie. *Description topographique, physique, civile, politique et historique de la partie Française de L'isle Saint-Domingue.* New edition, 3 volumes, Paris, 1958.

Murray, Gerald F. "The Evolution of Haitian Peasant Land Tenure: A Case Study in Agrarian Adaptation to Population Growth", PhD thesis, Columbia University, New York, 1977.

Myrdal, Gunnar. *Asian Drama. An Inquiry into the Poverty of Nations.* New York, 1968.

Nicholls, David. "Rural Protest and Peasant Revolt in Haiti (1804–1869)", in Malcolm Cross and Arnaud Marks (eds). *Peasants, Plantations and Rural Communities in the Caribbean.* Guilford and Leiden, 1979.

Palmer, Ernest Charles. "Land Use and Landscape Change along the Dominican-Haitian Border", PhD thesis, University of Florida, Gainesville, 1976.

Pierre-Charles, Gérard. *Haiti: Radiografia de una dictadura – Haití bajo el régimen del doctor Duvalier.* México, D.F., 1969.

Schmidt, Hans, *The United States Occupation of Haiti, 1915–1934.* New Brunswick, 1971.

Zuvekas, Clarence, Jr. *Agricultural Development in Haiti. An Assessment of Sector Problems, Policies, and Prospects Under Conditions of Severe Soil Erosion. Mimeo, US/AID, Washington D.C., 1978.*

# Resource Allocation and Entrepreneurship

BY JAMES M. BUCHANAN*

*The critique developed in this paper involves the neglect of the entrepreneurship role in the theory of the resource allocation process of the economy. My critique is similar to and related to that advanced by Israel Kirzner in his book, Competition and Entrepreneurship (Chicago, 1974). It also has close affinities with the classic work of Joseph Schumpeter, whose book, Theory of Economic Development, was published in its first version as early as 1911.*

## I. An Elementary Statement of the Coase Theorem

Persons trade when the relative evaluations of the units traded differ. When trade ceases, the traded units are held by those persons who place relatively higher value on such units than other persons. When there are no impediments to trade, all valued units are allocated to their highest valued uses, with values being settled at the moment of trade.

The elementary statements made above can be interpreted as one version of the now-classic Coase Theorem on the allocation of resources. But, as stated here, the question immediately suggested is: Why would anyone have thought differently? We can, I think, point to some reasons for the modern oversight of the quite elementary propositions that the Coase Theorem embodies. If we look, not at the *trading process*, but at the results or end-states of trade, and, further, if we look at the characteristics of equilibrium end-states, and implicitly make the assumption that all traded items are divisible into small units, the elementary statements made above do not hold. In equilibrium, all persons place the *same* relative evaluation on any unit of any traded item or commodity. Conversely, separate units have the *same* value in each use. The possible differential evaluations placed on inframarginal units of potential use become irrelevant in equilibrium adjustment. The relative evaluations placed on the inframarginal units will, of course, determine where the margins are located; that is, the evaluations over the inframarginal ranges will determine the final allocation of the total stock of any given traded item among separate traders, or among all pro-

jected uses or employments. In the simplest of trading examples, that of two-person trade in two-goods (A and B trade apples and oranges), with given initial endowments, we cannot predict how many apples and how many oranges each person will have acquired when equilibrium is reached until and unless we know something about the evaluations over the inframarginal ranges of anticipated usage. We can say, however, that for the quantity of any item purchased in unimpeded trade, the purchaser will place a higher relative value on such quantity than anyone else. This statement holds even if, for a marginal unit (and any unit if marginal), relative valuations are identical as over all persons.

To this point, I have done nothing more than restate some of the most elementary principles of economics. I want, however, to use this restatement for a purpose. I want to examine some of the implications of the proposition that the stock of traded items is allocated among all potential purchasers or users so as to insure possession or ownership by those of the group who place the relatively highest value on all units of the stock. To my knowledge, the implications of this elementary proposition have not been fully explored.

## II. The Allocation of Final Goods

Initially and by way of introduction, we may restrict analysis to trade in final goods, or consumption end-items. We can say, almost tautologically, that unimpeded trade will generate an allocation of a fixed stock of such a good among persons such that "utility", evaluated in some numeraire good, is higher than that achieved in any other

conceivable allocation. Since, however, "utility" is not interpersonally comparable, this statement really adds nothing at all to understanding. There is no way of getting at the question: Why does Mr. A value the bundle of oranges that he retains in full-trading equilibrium more than Mr. B? To answer such a question, we should have to get "inside" Mr. A's utility function itself, something that economists have been reluctant to attempt.

Until and unless we could begin to answer such questions, however, there is no predictive content in economic theory at this level. By our unwillingness and/or inability to measure "utility", we insure that there is no empirical, "objective" content in analysis. There is no way that the economist can lay down presumably objective conditions or standards, which might be empirically checked, in order to guarantee efficiency (highest value) in the use of a final good. By his own methodological constraints, the economist is forced to search for his efficiency criteria by an examination of the trading process rather than by any examination or testing of the end results. The economic theory of the exchange economy, with initial endowments of final goods, must be beyond the pale for "positive economics" in the modern sense.

### III. The Allocation of Intermediate Goods

The initial consideration of the proposition with respect to trade in final goods is useful for purposes of contrast and comparison with the implications for intermediate goods, those goods, resources, assets, or services that do not enter directly as end-items in the individual's utility function. These goods are traded, but they (or their services) are not consumed directly (transformed into utility). These goods are indirectly consumed via a productive process that involves their transformation into final goods and, through time, into "utility".

The characteristics of the trading process are unchanged. Such items or goods are allocated to their most highly valued uses to the extent that trade is unimpeded. But, if these items do not yield direct utility to their purchasers-users, why should they be valued *differently* by different persons? Complexities arise at this point if we remain within the certainty paradigm of much modern theory. If the "capacity to produce" is something inherent in a unit of intermediate good, and if this "capacity" is known with measurably objective certainty, and by all persons, it follows that

all persons will value such a unit at precisely the same amount. In this case, we should observe no trade to take place. To rationalize or to "explain" trade in nonfinal or intermediate goods, therefore, we must introduce differences among potential traders in their subjective assessments of the potential "capacity" embodied in units of such goods. In general, a purchaser does not buy a fork-lift truck because he has a "taste" for this equipment. The prospective buyer must somehow think that a unit of an intermediate good or service has a higher capacity to produce final goods, and hence utility, in some ultimate sense, than does the person who might enter as a prospective seller on the other side of a trade.

Let me go through some elementary economics by way of getting to some of my main argument. Consider an example, that of a potential owner-operator of a warehouse, who enters the market for fork-lift trucks. In the familiar diagram of Figure 1, we can depict his "demand" for units of the good, and, given the fixed supply price,, he will purchase, illustratively, seventeen units. This quantity will maximize the purchaser's rental value of the complementary resource inputs (labor, pallets, space), as shown by the "buyer's surplus" triangle, S. We can think of this same owner-operator as entering the market for each one of these complementary inputs, and we might depict his surplus-maximizing solution in the same manner as that shown for fork-lift trucks in Figure 1. We can model his decision process as one of simultaneous determination of the surplus-maximizing rates of purchase (or hire or lease) in all of the input markets.
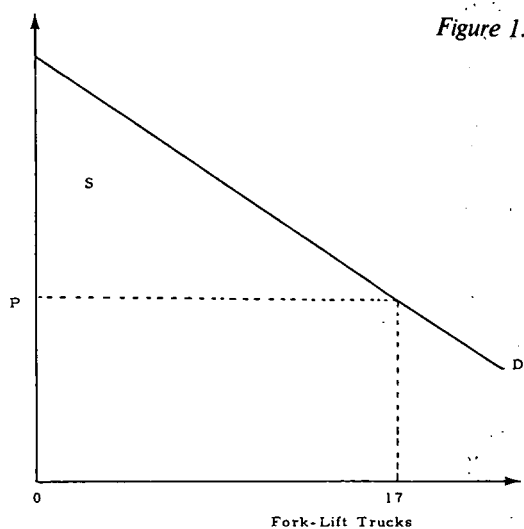


*Figure 1.*

Fork-Lift Trucks

In full competitive equilibrium, the payment for all inputs will just equal revenues derived from the sale of the goods or services produced; there will be no economic profits. In his decisions, however, the owner-operator of the warehouse must seek, and he must expect to find, positive profits. He will try to maximize net surplus or net rents. If he has no anticipation of securing profits, over and beyond the required outlays on the resource inputs purchased, he will not, of course, organize production. That is, he will not "trade" with the suppliers of these inputs, since, by assumption, he has no "taste" for the inputs, as such.

## IV. The Circulation or Evenly-Rotating Equilibrium in the Stationary Economy

I get confused, however, when I try to think out the full implications of this elementary account of the behavior of the entrepreneur in a competitive environment. In full competitive equilibrium, as noted, we cannot allow for returns to pure entrepreneurship over and beyond the opportunity costs of the resource inputs actually used. But does this fact alone not suggest that an equilibrium becomes *logically* impossible? There seems to be nothing in the system to "make the wheels go round," so to speak. Schumpeter speaks about the circular flow in full stationary equilibrium, where all economic agents find that their expectations are fulfilled, and where they repeat the same behavior period after period, given no change in the exogenous parameters of the system (wants, resources, technology).

I can model such a circular flow process, the evenly-rotating stationary state, under the assumptions of a pure exchange economy, in which each participant commences with an endowment of end-products, the same each period and received in some "manna-from-heaven" distribution. In such a setting, each person would repeat the same behavior in each period; he would trade the same units of his initial endowment for the more highly valued end-items in order to maximize his utility. Failure to behave in this fashion would mean lower utility attainment. The same prices will be reestablished in each period; the same final allocation of goods to persons will be consumed.

In a similar way, I have no difficulty with a production economy when each person is assumed to be endowed with a capacity to produce a single end-item, and where the use of this capacity is also an argument (a bad) in each person's

utility funtion. Nor is there any difficulty in relaxing this restriction to allow that productive capacity may be used in producing several consumption goods, so long as we describe the utility function to include arguments for each use of the inputs. In both of these models of a production economy, the individual's utility maximization behavior will, just as in the pure exchange economy, lead to the same allocation of capacities in each period, the same set of prices, the same final allocation of consumption among persons.

The reason for the constant repetition of the equilibrium allocation, period after period, is found in the fact that each person, by behaving any differently, will be in a worse position. There must be differential advantages to be gained from behaving so as to generate the repeatable equilibrium solution, even if these advantages be infinitesmal at the appropriate behavioral margins.

Consider the case of a person who can produce either gidgets or widgets with his talents. Why would he spend two hours on gidgets and six on widgets each and every day? He would do so only because any different behavior would reduce his utility. Hence, producers' surplus serves the selfsame allocation purpose as consumers' surplus in the allocation of final end-items.

As normally stated, however, producers' surplus, scarcity rents, profits, are not supposed present in the abstracted general equilibrium of the fully competitive economy. Owners of resource inputs are presumed to be confronted with alternative employments, each one of which yields the same return, and, further, these resource owners are presumed to be indifferent as among the separate potential uses. In such a setting, however, why will the equilibrium allocation be repeated period-by-period? Clearly, there is nothing unique in the solution if rents are wholly absent, even for a single, solitary unit of input.

I offer no answer to my own puzzle here. I leave this to the so-called "economic theorists", but you can see how this puzzle relates directly to my interest in entrepreneurship and its role in the allocative process. If rents or profits are allowed as possible, or even if they are only thought to be possible, entrepreneurial activity will "drive the system", and, of course, competitive entry will always put pressure on observed profits and rents such as to erode these. I ask only whether or not we have modelled an internally contradictory structure that leaves no room for producers' surplus, profits, or rents, and

whether or not such a modelling has inadvertently modified the mind-set of the economists who then come to think of idealized allocations without allocators.

Should we drop the very notion of general equilibrium, even as a logical construction, once we recognize the contradiction? Or should we try to reconstruct it so as to allow universalized producers' surplus? Note that if we allow entrepreneurs in the model, we do get a solution to the allocation problem, *ex ante*. All resource units flow necessarily to the most highly valued uses, as viewed by the entrepreneurs who implement and carry forward the input combinations and who organize production.

## V. Optimistic Entrepreneurs

The potential producer, the entrepreneur, must be relatively "optimistic" about his ability to combine resources so as to achieve a positive rent or surplus. He places a higher expected value on the bundle of resource inputs that he decides to purchase than anyone else in the economy or trading network. What does this statement imply about realized values, about realized rents or surplus, realized profits? As noted above, at the moment of entrepreneurial choice, expected returns, expected profits, must be positive. In competitive equilibrium, however, profits will disappear. To the extent that unimpeded trade, including freedom of entry and exit into and from all markets, generates any adjustment toward equilibrium, even if such a state is never attained, realized rents will tend to fall below expected profits. From this it follows that the entrepreneurs, as a group, or in some representative sense, must be *disappointed*. This result must hold despite the possible presence of individual cases in which realized rents might exceed expected rents. Realized rents or profits may range the spectrum from higher-than-expected levels at the one extreme to large negative values at the other. Entrepreneurs in the first group, that is, those whose initial optimism pays off beyond expectations, need not be disappointed. But, by necessity, these entrepreneurs represent only the tail of the distribution that we may assume to be symmetric in some fashion.

The tendency of the market process to insure that resources come into the usage and ownership of those persons who are most optimistic about their productivity, who place the relatively highest value of these resources *ex ante*, is, at the same time, reflected in the mirror image of ubiquitous entrepreneurial disappointment. Plans are not realized, and, on average, rents fall short of those anticipated. The disappointment of entrepreneurs has several important implications. Because plans do not live up to expectations, entrepreneurs will be led to turn over assets, to modify their projects, to change their rates of purchase of resource units, more frequently than that rate that might be predicted under the standard assumptions that are imbedded in economic theory. Under the latter assumptions, error leading to negative profits may be made, leading to corrective adjustments on the part of entrepreneurs. When the generality of entrepreneurial disappointment is reckoned on, however, it becomes clear that entrepreneurs will tend to modify plans and to shift among separate projects even when realized profits may be positive, possibly strongly so.

For illustration, consider two separate projects undertaken by two separate entrepreneurs, projects that are not directly related, one to the other. Each of the two entrepreneurs expects to secure, say, $ 1000 in net profits when he makes the decision to organize production and proceeds to purchase the inputs required. (Note that we cannot define expected profits in terms of a "rate of return" on anything.) Both are typical or representative entrepreneurs, and, hence, are disappointed to find that realized profits or surplus amounts to only $ 500 in each, each still positive but not so high as anticipated. Each entrepreneur, viewing the alternative project to his own, may now consider switching his efforts, despite the presence of an observed profit level that is no greater than that realized. It is quite possible to get such a switching among projects without any change in the total of profits realized. This result could not emerge under the standard assumptions of economic theory, which would suggest, in this example, that both of the entrepreneurs would remain in production of the projects initially commenced.

This tendency to shift resource combinations, to change projects, will, of course, be dampened to the extent that specificity is a necessary component of project choice. If an entrepreneur, in organizing production for an initial project, finds it advantageous to convert transformable units of resources into forms that are specific to the project, the differential between realized quasi-rents and the scrap or disposal values of assets may seriously inhibit the switching of production or production technique.

## VI. Managerial Rotation[1]

One direct implication of this analysis of entrepreneurial disappointment involves the rate of turnover or rotation of managers of enterprises. Entrepreneurs hire managers to supervise production; managers are among the resource units purchased. But, by definition, managers "manage"; that is, they are expected to exercise discretion in the supervision of other input usage. Managerial talents tend to be readily transferable as among alternative employments. These qualities of management, combined with the ubiquity of entrepreneurial disappointment, suggest that managers will be quite vulnerable to discharge and rotation, and quite independently of any problem in the internal incentive structure that may affect the behavior of managers themselves. In other terms, the effect described here would occur even in the extreme case where managers behaved, in each and every instance, as if their own interest should be identical with that of the entrepreneur.

Professional sports enterprises offer an exellent real world illustration of the argument here. Owner-entrepreneurs are optimistic about the prospects of franchises, and they tend to assign expected values to franchises in excess of any values that might be realized. They hire managers, and they tend to be disappointed with managerial performance, quite independently of any measure of the actual quality of managerial decisions. Frustrated when results do not match up to anticipations, owners fire managers and hire new ones, for the most part from the pool of available persons who have, themselves, been fired by other owner-entrepreneurs and for the same reason. We observe a high rate of managerial turnover without noticeable changes in the relative fortunes of the separate franchises.

## VII. Entrepreneurship and Risk-Taking

Entrepreneurs act because they expect to make profits. Their action has no relationship to the bearing of risk or uncertainty, as such. An entrepreneur may exhibit risk preference, risk neutrality, or risk averseness. To the extent that he is risk averse, the expected profit from any project necessary to bring him over the threshold of positive action will be greater than that which would be necessary under risk neutrality or risk preference. The risk-averse entrepreneur would never undertake a project that promises a marginally-positive expected surplus, if there is risk or uncertainty involved.

It may be argued, further, that since each entrepreneurial choice is unique, genuine uncertainty must be present. As Shackle has stressed, since the properties of the whole experiment, which may embody Knightian risk rather than Knightian uncertainty, cannot be relevant to the unique choice that must be made, entrepreneurs must choose among actions that are necessarily uncertain as to outcomes. This argument may be, indeed must be, granted, but there remains the possibility of arraying, at least conceptually, the entrepreneurial choice situations in terms of their uncertainty characteristics. In the one extreme, an entrepreneur may be "relatively certain" that the outcome he predicts will, in fact, occur consequent to his action. The entrepreneur who buys wheat in one market and sells it simultaneously in another, and for a different price, is acting under conditions of "relative certainty". When we examine entrepreneurial choice under conditions of relative certainty somewhat more carefully, we can show that profits, the residual rewards to entrepreneurship, are not properly described or defined as a reward for risk or uncertainty bearing on an economy. Consider the entrepreneur-arbitrageur who buys wheat in one market and sells it simultaneously in another. There is little or no risk or uncertainty involved, and such an entrepreneur may be higly risk averse. The profit that he makes rewards him for his ability to "see" the profit opportunity and to act upon it. He is rewarded for "creating" value by sensing the differentials in price. Without such a prospect of reward the value would not, in fact, exist or come into being at all.

In most conditions for entrepreneurial choice, of course, decisions must be made under uncertainty, and entrepreneurs must, in one sense, accept such uncertainty bearing as a necessary characteristic of their choice situation. But a willingness to bear uncertainty is surely not a sufficient condition for entrepreneurship. There may exist many persons, who are genuinely risk-loving, and who will gladly take on the uncertainty of investments in projects that are presented to them. Such persons may, however, wholly lack any ability to see profit opportunities, to invent in their mind's eye new arrangements, new technology, new resource combinations. There may be no correlation at all between personal talents in this respect and personal proclivities to take risks.

## VIII. Entrepreneurship and Time

To this point, I have done little more than recast slightly, and with the somewhat interesting managerial implications, the theory of entrepreneurship presented by Kirzner. I want now, however, to diverge from Kirzner's conception in one important respect, namely in his emphasis on the absence of any necessary relationship between ownership and entrepreneurship. I can appreciate Kirzner's purpose; he sought to divorce or to separate the economic function or role played by the owners of capital assets from the role or function of the entrepreneur. I have no quarrel with such separation, which is essential for logical clarity. As Kirzner emphasized, the pure entrepreneur need hold no assets at all. His idealization is the instantaneous arbitrageur, who simultaneously enters separate markets on differing sides, seeking profit in the process.

This idealization is a biased one, however, and is best described as an extreme end of a possible spectrum of models for entrepreneurship, and in no way "representative"of the sort of entrepreneurs Kirzner seeks to place in his motivating roles in a competitive economy. Almost universally, entrepreneurs seek their profits by holding, or owning, assets *through time* rather than the instantaneous arbitraging modelled by Kirzner. I do not suggest that they hold capital assets "as capitalists", that is, in order to secure a rate of return of the ordinary sort. Quite the contrary, and Kirzner is quite correct in stressing the difference here. My point is rather that, in order to engage in entrepreneurial arbitrage, defined in the large, most "traders" must work in time. They do not "hedge" as if they are the classic-case flour millers. Most entrepreneurs buy in one market *now*, and expect to sell in another market *later*, or vice versa, or at least I should argue that this is a more representative model of entrepreneurial activity than Kirzner's instantaneous or simultaneous model.

This model suggests that confusion about the pure entrepreneurial role is especially likely to emerge, since the temporal aspects suggest the risk or uncertainty-bearing function previously discussed as well as the capitalist or pure ownership function associated with the productivity of capital itself. The pure entrepreneur, however, sublimates as inessential or inconsequential *both* the risk-bearing and the ownership role, which he may, nonetheless, be required to occupy in order to take advantage of the profitable oppor-

tunity that he thinks he sees before him. The pure entrepreneur may, of course, borrow sufficient funds to finance the required outlay on the assets to be transferred to his ownership, and the rates at which he borrows may be even higher than any nominally-computed "return" on the value of these assets. In this setting, the entrepreneur is not at all a "capitalist" in any net-asset or net-wealth sense. He may be, and probably is, more normally in a net debtor than a net creditor position. But nonetheless, the entrepreneur must secure, and hold, title to the particular asset, or asset bundle, that he purchases in order to secure for himself the anticipated profits from later resale at a higher price. This "arbitrage through time"model of entrepreneurship can incorporate examples extending from ordinary speculation in real estate through the organization of production of final goods and services.

## IX. Entrepreneurship and Inflation

Unless the temporal setting within which entrepreneurial action takes place is recognized, the effects of anticipated inflation upon entrepreneurship, and, through this, on the dynamics of the economic process, tends to be obscured. Implicitly, or by presumption, the role of the entrepreneur, and of entrepreneurship, discussed in preceding sections of this paper is carried out in the context of an economy described by monetary stability, at least within limits of tolerance. That is to say, the individual entrepreneur, who purchases resource units for the purpose of exploiting a profit opportunity that is not universally seen by all participants in the economy, acts in the expectation of being able to create *real value*, as measured by the response of market participants when confronted with the opportunities that he constructs. Although individual entrepreneurs are not conscious of such unintended consequences, their action, in net, is generative of increases in real product value in the economy. Resources are reallocated via entrepreneurial creativity in such fashion as to increase overall value productivity.

Let us superimpose upon this dynamic model of economic process governmentally-generated, continous, and anticipated inflation. The predicted effects are clear. Opportunities for entrepreneurial profits emerge that do not necessarily generate increases in real value. Anticipated inflation opens up generalized opportunities for arbitrage-through-time rather than the specialized opportunities open to ordinary entrepreneurship in con-

ditions of monetary stability. Attempted exploitation of the generalized opportunities here will drive up the prices of real-valued assets that are durable relative to prices of either nondurable goods or of claims to nominal-valued assets. The specialized opportunites for genuinely creative entrepreneurship will, of course, continue to exist, potentially, in the inflationary setting. However, exploitation of these opportunities is made more difficult by the relative bias introduced in the structure of temporally-designated prices of goods. To take advantage of a genuine opportunity, an entrepreneur must secure title to real-valued assets. In order to do so, he must assume a fixed-value liability; he must issue "bonds", denominated in nominal yields. If the purchaser of such "bonds" (the seller of the real asset) and the entrepreneur place the same expected value on the anticipated rate of inflation, and, further, if this rate is universally expected by *all* persons in the economy, the effects that have been suggested here need not arise.

If, however, we allow the more plausible realistic model in which *some* but not *all* persons in the economy fully anticipate the inflation, the relative price bias note must emerge. This point is worth developing in some detail.

Assume that the government is committed to maintain a specific rate of continuing inflation. Assume, further, that the full effects of this policy are predicted by only *some* of the economy's participants (facetiously, we may call these "the economists"). Members of the latter group will see the arbitrage opportunities available to them as a result of their superior knowledge about the effects of the government's announced behaviour. These persons (the "new entrepreneurs") will, therefore, reduce current holdings of money and nominal claims ("bonds") and increase current holdings of durable assets that are expected to appreciate in value over time. Prices of the latter assets will rise; prices of "bonds" will fall.

There is, however, no necessary intersection between the set of "true entrepreneurs" and that set of "new entrepreneurs" attracted to invest in real assets solely because-of the anticipated inflation. For members of the former set who are not, simultaneously, members of the latter set, the terms-of-trade will have shifted dramatically against them. Potential profit opportunities which might exist in monetary stability vanish in the *ex ante* sense, and no entrepreneurial reallocation of resource toward generating higher real values takes place. The dynamic development of the economy is dampened.

The effect on economic process generated by this dampening of entrepreneurial acitivity is not inconsistent with *ex post* findings that, as a group, entrepreneurs secure relative gains from inflation. We get the somewhat paradoxical result that while inflation may substantially reduce the number of entrepreneurial projects, among those introduced there will be relatively few failures or bankruptcies. Entrepreneurs may gain, *ex post*, from inflation due largely to the net monetary debtor status or, what amounts to about the same thing, to entrepreneurs' necessary role in the temporal arbitrage process. To the extent, however, that the set of "new entrepreneurs", created by the inflationary expectations, and the "true entrepreneurs" do not match precisely, inflation must have the effect of preventing some resources from those employments or usages where they are most highly valued, in the *ex ante* sense. A land parcel held for potential inflationary gains by a "new entrepreneur" may not be worth the asking price to a "true entrepreneur", who may envisage genuine development prospects but who may not fully anticipate the inflationary effects of government policy. As a result, the economy produces lower real value than otherwise might have been produced.

Resources flow, via the activity of entrepreneurs, to those uses that promise the highest value, as estimated by entrepreneurs at the moment of market exchange. The introduction of inflation does not modify this basic proposition. But inflation does have the effect of distorting the prospective values estimated by entrepreneurs. In this context, it is worth keeping in mind that the "highest valued uses" do not exist independently of entrepreneurial estimates. "Highest valued uses" for resources are "created" in the imagination of entrepreneurs, and any distortion introduced into entrepreneurial estimates may destroy potential value never to be replaced.

As noted earlier, the effects of anticipated inflation discussed here take place only so long as some potential entrepreneurs fail to incorporate the correct inflationary anticipations in their own estimates. If and when *all* potential entrepreneurs come to act upon the same anticipated rate of inflation, along with all other participants in the economy, the distortions will, of course, disappear. Critical errors may be made, however, in the failure to distinguish between an inflation anticipated by *some* persons in the economy and an inflation anticipated by *all* persons in the econ-

omy. The state of "equilibrium expectations" describing the latter situation may not be reached until the end of an extremely long temporal sequence.

## X. Conclusions

This paper has developed no central "theme" or "principle". It should perhaps have been entitled, "Notes on Entrepreneurship". In writing this paper, in two versions separated by several months, I have had the feeling that many other implications than those discussed here would emerge. To this point, no others have dramatically appeared. I become more convinced, however, that a "breaking out" of the intellectual constraints imposed on so many of us by the equilibrium constructions of neoclassical economic theory is necessary if we are to understand the economic process properly, and through some such understanding, begin to get some handles on how the dynamic potential of the market order might, once again, be harnessed.

## Footnotes

1 The discussion in this section is due to a suggestion by Roger Faith.

# Power and Profit in Hierarchical Organisations

BY DENNIS C MUELLER*

The development of economics as a social science has rested on two postulates concerning human behavior: individuals act out of self interest, and are rational. Thus, consumers are assumed to maximize their utility; entrepreneurs maximize profit. The economy is driven by self interest, the corporation by the profit motive.

Recently the postulates of rationality and self interest have been extended to the study of political science through the development of the public choice area. In one of the pioneering works in this area Anthony Downs (1957) postulated that candidates pursued their self interest by trying to maximize the number of votes they obtained. Although vote maximization has proven to be a plausible and useful assumption to explain candidate behavior, it clearly cannot explain all political behavior, since many politicans are not elected and/or cannot be re-elected. In particular, this postulate cannot explain bureaucratic behavior.

The classic analysis of bureaucracy is, of course, Max Weber's (1947) and the natural objective for the bureaucrat, following Weber, would be power. The large corporation is run by managers seeking profit; the public bureaucracies by individuals hunting for power. Economic man pursues profit; political man power.

In the pages which follow I attempt to develop an analogy between power and profit, and use it to analyse the objectives and conflicts that arise in hierarchical organizations. Since the literatures on both of these subjects are long and tortuous, I make only selective reference to each. Let us begin with power.

## I. The Concept of Power

At the most intuitive level the word "power" connotes the ability or capacity to do something, see Wagner (1969, pp 3–4). But "something" can stand for a variety of objects, each of which leads to a different conception of power. Physical power is the ability to apply force. Economic power is the capacity to purchase goods, and so on. Political power must be defined as the ability to achieve certain ends through a political process. In this essay we shall take a rather broad view of the latter, considering virtually any collective decision-making body or organization from a committee to a bureaucracy as being governed by some form of political process. To observe the exertion of political power it is necessary that at least some participants in the political process have conflicting goals. If all members of a committee favor the same alternative as A and this alternative is chosen we cannot say that $A$ has exercised power. If only $A$ favors an alternative and it is chosen, $A$ has political power.

Political power can arise directly from the rules by which the political process operates. These rules might simply grant $A$ a dictatorial right. Under most rules, the committee chairman has more capacity to influence the outcome than other members, yet he need not be the most powerful member of the committe. What interests us here is not the direct capacity to influence an outcome granted by the rules, but the differing capacities individuals have to influence a collective decision, independent of the set of rules.

Bertrand Russell (1938) listed three ways in which an individual can exert influence in a political context (1) by direct physical power, e.g. imprisonment or death, (2) by offering rewards and punishments, and (3) by exerting influence on opinion through the use of education and propaganda. The first two are obviously closely related to procedural power. The dictator may have authority to imprison or execute subordinates, they most certainly will not have similar legal authority over him. As Cartwright (1965, p 139) has observed, "Of the many possible means of influence, persuasion is commonly advocated as most suited to a democratic, or rational, social system". Thus, the third of Russell's sources of influence is of most interest to us here. On the surface, it also

seems to come closest to our description of power. For education, propaganda, and persuasion are all forms of information. As we shall attempt to demonstrate, political power, other than of a procedural kind, is possessed by those who have information. Uncertainty creates the potential to exercise power, information provides the capacity to do so.

Although information will provide the most power in a political process governed by persuasion, it is not limited to these most democratic forms of political interaction. To illustrate the generality of the uncertainty-information-power nexus we first examine a situation that seems to come closest to Russell's first source of influence, pure physical power. Consider the classic power struggle encapsulated by the demand "your money or your life". $G$ has a loaded gun which he aims at $W$ and demands that $W$ give his loaded wallet over to $G$. Here we have what appears to be the simplest case of power by force with information playing no visible role. Let us examine more closely. $W$ must choose whether to hand over his wallet or not. He must, therefore, predict what $G$ will do should he not hand it over and if he does. Suppose $W$ *knows* that $G$ will not shoot in either event, $G$ is then without any power. $W$ keeps his wallet and $G$ does not shoot. Suppose $W$ knows $G$ will shoot in either event. Again $G$ is without power, i.e. the ability to command, since $W$ knows the wallet now belongs to $G$, and it is simply up to $W$ to decide whether he wants to give it to $G$ and then be shot, or let him take it after $W$ is dead. The same holds true for the case when $G$ will shoot if he does not get the wallet, but will not shoot if he gets it. If $W$ knows this with certainty, suppose $G$ is a programmed robot, $G$ is without any real power to command. The choice is $W$'s, whether to live without his wallet or die with it, and $G$ merely carries out his programmed action following the real decision by $W$. The only situation in which $G$ can actually command $W$ to do something against his will, is when $W$ does not know what $G$ will do following $W$'s action. $W$ might then give $G$ the wallet when $G$ would not have shot him anyway.

It is in this situation, and really only in this situation, that $G$ can be said to be exercising political power over $W$, as political power is typically defined, see Dahl (1957, p 80), Simon (1953). If $G$ would not shoot $W$ if $W$ failed to give him the wallet, and yet he can get $W$ to give him the wallet, he has succeeded to get $W$ to do something he would not otherwise have done. $G$ has done so,

however, not solely because he has a gun, but because $W$ is uncertain about what $G$ will do with the gun. It is not the presence of the gun *per se*, but the uncertainty that accompanies it that gives $G$ power. In the absence of the gun, $G$ does not have power over $W$ because $W$ is not worried that $G$ will kill him. If $G$ gets a gun, he will have power over $W$, because, or more precisely if and only if, $W$ is uncertain about what $G$ will do with it. $G$ has power because he has the information about what he will do and $W$ does not.

In this example, the gun plays the role of procedural power and clearly it places $G$ in a better position to achieve his goals than $W$. But it alone does not determine the outcome so long as there is uncertainty on the part of the individual to the other's reactions. It is this uncertainty that gives $G$ power over $W$, and can give $W$ some power over $G$.

The importance of uncertainty and information can be further demonstrated by slightly changing the example. Suppose that $W$ has buried the wallet someplace in his yard and only he knows the location. Now there is considerable uncertainty on both sides: $W$ not knowing whether he will get shot, $G$ not knowing the location of the wallet. Given the increase in uncertainty and relevant information in the hands of $W$ his power should be enhanced. He can now quite possibly force $G$ to unload or throw away his gun in exchange for information on the wallet's location. Indeed, he might get off with both his life and his wallet. Even though the advantage of force still lies on the side of $G$, the increase in $W$'s possession of relevant information gives him the potential for exercising considerably more power over $G$.

As a final extreme example, assume $G$ and $W$ both have wallets and known programmed response patterns in the event that one has a gun. A gun is given to one on a flip of a coin. Given the programmed reactions of each, no real power is meted out via the coin flip, although the flip will affect the lives and/or wealth of $W$ and $G$. What power that exists in the situation is with the coin flipper, or a fate which knows the outcome of the flip.

Returning to Russell's list of sources of power, we can see that it is the uncertainty that surrounds a dictator's use of physical power, or a supervisor's issuance of rewards and punishments that allows them to control their subordinates. If $B$ knows with certainty that $A$ will give him a reward if $B$ does $X$, the rules require it, then $B$ in carrying out $X$ exercises as much power over $A$ as $A$ does

over *B*. In a bureaucracy in which no uncertainty existed, lines of authority might exist, but no real power would accompany authority. All employees would know all of the possible events that might occur and all could predict the eventual outcomes or decisions that would follow each. Employee grievance procedures would be completely codified and both the supervisor's and the employee's reaction to any situation would be perfectly predictable. In a world of complete certainty, all individuals are essentially acting out a part, "going by the rules", and those at the top of the bureaucracies are as devoid of discretionary power as those at the bottom. All power is purely procedural, see Simon (1953, p 72).

This type of situation comes close to the conditions existing in the French Monopoly Michel Crozier (1964) described in *The Bureaucratic Phenomenon*. As Crozier depicts it the monopoly does operate in a world of certainty – with one exception – the machines sometimes break down. This places the women operating the machines completely under the power of the mechanics responsible for repairing them, since the women have a quota of output for each day and must work harder to make up for any down time. More interestingly, the supervisors who nominally have more authority also have less power than the mechanics. Since the mechanics know how to repair the machines, and the supervisors do not, the supervisors are unable to exert any real control over the mechanics, see Crozier (1964, pp 98–111).

It is instructive to note the tactics used by the mechanics to preserve their power. The operators were severely scolded for "tinkering" with their machines in an effort to keep them going or repair them. Only the mechanics knew how to repair the machines; each machine was different and just how it needed to be fixed was known only to the mechanics; repairing them was an art not a science. When clashes arose between the mechanics and the supervisors it was over whether the latter could, on occasion, work at repairing the machines. The supervisors were further hampered in this endeavor by the continual "mysterious" disappearance of machine blue prints from the factory. The mechanics always worked without the aid of blue prints.

It is easy to extend Crozier's description of the tactics employed by the mechanics to maintain their control of information and power to other groups of experts. One of the first things any group does to protect its position is to develop a set of terms or jargon that makes much of what it does inaccessible to outsiders. This can be further buttressed by perfecting techniques of analysis so complicated that outsiders cannot follow them. This done it becomes extremely difficult for those outside the group to take away or evaluate the information possessed by the expert. Examples of this behavior are obvious. Scientists and engineers perhaps come first to mind. In these professions the nonspecialist is clearly at a loss to understand and exercise effective control over the professions. Even within the disciplines the tendency is for information boundaries to arise giving groups power vis-á-vis their colleagues. Thus, the inability of one branch of physics to evaluate the work of another strengthens the position of the inaccessible branch in gathering R&D funds, grants, department positions, or what have you on the basis of its own criteria. The "pecking order" both across and within disciplines tends to be from "hard" to "soft" science on the grounds that the more theoretical or mathematically oriented hard scientists can or could always understand and evaluate the "soft stuff", while the reverse is not necessarily true. The counter argument by the more applied is a rather weak claim of expertise because the theorists are not really familiar with the data or the institutions.

Other professions attempt to create and maintain power in the same way. Consider law. Here is a profession whose language could be, and once was, accessible to the average citizen. Over time, however, the profession has so complicated the language and procedures used in the judicial system that it is nearly impossible for an outsider to participate without hiring a lawyer. The medical profession follows a similar strategy, with the practice of writing prescriptions in Latin being an interesting illustration.

Lacking Latin or mathematics to conceal information and preserve their power, individuals typically resort to the more blatant device of secrecy. Examples ranging from the fraternal "secret handshake" to the classification procedures of the Pentagon and CIA come easily to mind. In each case the purpose is the same, to protect the insider's position by keeping relevant information from the outsider. Although the purported purpose for classifying many documents is to preserve national security by keeping them out of the hands of the nation's enemies, the true, intended "outsiders" often appear to be our own citizens, and the "insiders" whose security is being protected, government bureaucrats.[1]

Crozier further butresses the hypothesis that

uncertainty is the source of power by examining the seemingly anomalous preference for technological change by the director of the Industrial Monopoly and the resistence to this change by the technical engineers. On the basis of social background and status the technical engineers should be more liberal and promote technological change, while the more conservative backgrounds of the directors should lead them to resist it. Decisions to institute changes in technique are made by the directors, however. In the absence of these changes, decisions are sufficiently routine that effective control lies with the engineers. Thus, the only time that the directors can effectively demonstrate their authority is when they initiate changes in plant technique. Uncertainty is then introduced, with the top directors in possession of the relevant information on the new technique. Following the change, uncertainty gradually diminishes, routine returns, and power passes down to the lower levels, until the directors are forced to introduce another change in technique, see Crozier (1964, pp 155–5).

Again, one can easily think of additional illustrations of the importance of information in establishing a group's power. Perhaps, the best one is that of the military. Here one has a situation in which uncertainty, over a weapon's effectiveness, levels of preparation, offensive and defensive strategies, etc is endemic to the activity. This gives the military a strong advantage over other federal bureaucracies in obtaining funds from both Congress and the Executive Branch. The development of an impenetrateable jargon, classification of data and so forth, all serve to maintain this uncertainty and strengthen the power of the leadership of the military hierarchy who have or claim to have the relevant information. The otherwise surprising preference of one of the oldest and most conservative bureaucracies, the military, for new and more sophisticated weapons systems, becomes understandable by analogy with the case studied by Crozier.[2]

These examples hopefully illustrate the role uncertainty and information play in creating and distributing power in bureaucracy. We shall return to an examination of bureaucratic power, after investigating the role information and uncertainty play in generating economic profit.

## II. Profit

Consider a world of perfect certainty. All tastes and technologies are known. Labor, land and cap-

ital are combined to produce goods and services. Competition ensures that the prices on all goods and services are driven to the point where they just cover factor input costs. There is no residual left for the entrepreneur (other than a normal compensation for whatever labor services he provided), since there is nothing that requires entrepreneurial skill in a world of perfect certainty.

When uncertainty exists revenues and costs are not always equal. Unexpected changes in tastes, weather, competing technologies and so on produce changes in demand and cost schedules that leave positive or negative "residuals" between total revenues and costs. These revenues accrue to those who assume the responsibility for organizing the company, the entrepreneurs, and are defined as the profits of the firm.[3]

With uncertainty present, the possibility of "making" profits by correctly anticipating or inducing changes in tastes and technologies arises. The entrepreneur who knows what style of shoes will sell next spring, who knows that a certain technology will reduce costs, and so on, earns profits. Those who do not know these things or makes mistakes earn losses. Entrepreneurial activity thus consists of gathering and evaluating information on what will sell, and what will reduce costs. As long as one entrepreneur has information on what will sell, and others do not (are uncertain) he can earn a profit. Information on consumer tastes and innovations thus provides a firm with the ability to earn more than other firms in the market – with power over the market. This power dissipates as others acquire information about consumer tastes, and imitate the innovations. As uncertainty vanishes so do profits. The "perennial gale of destruction" described by Joseph Schumpeter thus consists of a process of gathering or creating new information which produces surpluses for those who have it, but soon is obtained by all, eliminating profits and setting the stage for a new finding, a new wave of profits and imitation, and so on.[4]

While Knight, Schumpeter and Coase stressed the importance of information not held by "outsiders" to the firm in generating profit residuals, more recently Alchian and Demsetz (1972), and Oliver Williamson (1975) have emphasized the importance of the distribution of information inside the firm to the generation and sharing of the residual. Alchian and Demsetz emphasize the team aspects of production within a firm. These can perhaps be best illustrated by considering pro-

duction a positive sum game of a prisoner's dilemma variety. The cooperative strategy can be interpreted as carrying out some previously agreed set of tasks at a given level of care and effort. The non-cooperative strategy is "shirking" on some of these tasks. All members of the team are better off if all adopt the cooperative strategy than if all do not, but some may still be tempted into shirking if they think they can do so without affecting the choice of strategy by other members of the team. In a small, productive team each member may be able to observe and monitor the behavior of the other members. In a large team this will be inefficient, however. A specialist at monitoring must be chosen. To ensure that this monitor does not, in turn, shirk he must be given the claim to the residual profit of the firm. Thus, in the Alchian-Demsetz theory, profit is also information-uncertainty related. The potential for profit exists in the behavorial uncertainties surrounding the prisoner's dilemma-teamwork production relationship. The profits accrue to the manager-monitors who gather information on other members of the team and ensure that they do not engage in shirking.

## III. Power, Profit and the Goals of the Organization

Uncertainty creates the potential for gains and losses, for correct decisions and mistakes. He who has the knowledge or information or intuition to make the correct decisions obtains power. This is true both within and outside of organizations. The individual who chooses the "right" career, buys the "right" piece of property, backs the "right" candidate, plants at the "right" time of year, and so on is ahead of those making the wrong choices. The general principle, then, is that uncertainty creates power for those having the information to make correct decisions in the face of the uncertainty. In the corporation, where the pursuit of profit is an accepted goal, this power is frequently monetarized in the form of high salaries, stock options, insider trading gains by the managers and so on.[5] In the nonprofit organization or the public bureaucracy power must more often be used to obtain nonpecuinary goals: security, leisure, status and prestige.[6]

The contrast should not be overdrawn, however. Corporate managers are interested in prestige, security and other nonpecuinary goals. And a number of writers have argued that the corporate manager's objectives are a package of pecuinary

and nonpecuinary goals rather than the maximization of profits.[7] Indeed, since reported profits are by custom and, to some extent, by law the property of the stockholders, managers must exercise their claim in the firm's residual in such a way so that it appears as a legitimate operating cost. Their options for doing so in a way that produces direct pecuinary benefits are limited, so that managers are almost forced to accept part of their share of profits in a nonpecuinary form. On the other side, there are a variety of possibilities by which public officials can gain financially from their position including the use of insider information, the receipt of gifts, bribes, kickbacks, etc.

The analogous role information and uncertainty play in the profit-oriented corporation and the nonprofit bureaucracy suggests that the behavior of individuals in these organizations, managers and bureaucrats, should in many ways be similar.[8] Both will seek to acquire information-power. Where they will differ, if at all, will be in how they utilize whatever power they possess to achieve their own personal pecuinary and nonpecuinary goals. We can thus expect managers and workers in industrial enterprises to adopt strategems similar to those Crozier describes in the two French bureaucracies of creating self-serving uncertainty, maintaining secrecy regarding information in one's possession, and so on.

The monitoring function managers serve in a teamwork organization suggests another strategy they might employ to increase their power. Recall that the need for monitor-specialists arises essentially because of the free-rider problem created by the prisoner's dilemma nature of teamwork production. The free-rider problem is worse, the larger the team. Thus, the need for monitor-managers and the importance of the information they possess will increase the larger the size of their organization.

Several writers have posited size or growth in size as goals of corporate managers, see Baumol (1967) and Marris (1964). The reasons given are typically the correlation between organizational size and managerial salaries, and the nonpecuinary rewards from managing a large, growing company. Our analysis suggests an additional reason why managers pursue size and growth. Growth can be expected to create uncertainty about the size of the residual profit and, thus, increase the value of the information managers gather. Increasing size worsens the free-rider problem, again increasing the value of the monitor-managers' information. In short, the power of managers within the cor-

poration should increase with size and growth. Managers should favor size and growth as corporate objectives, since they increase their power to achieve any other more direct personal goal the managers have.

The major constraint on management's claims on the profit-residual is the threat of outside takeover, see Marris (1964), Manne (1966) and Alchian and Demsetz (1972). The free-rider problem keeps the average stockholder from carefully monitoring managers, but the voting rights which accompany common shares provide incentives for outside entrepreneurs to buy out large blocks of shares and take over the company transferring the incumbent management's share of profit to itself. To do so, however, the potential takeover-raider needs to have information on the profit he can earn from a successful takeover. This is information that is possessed and for obvious reasons guarded by the incumbent managers. Here again size and, more specifically, diversification can increase the power of managers vis-à-vis potential takeover – raiders by increasing the volume and complexity of the information required to evaluate the potential gains from a takeover raid. Ajit Singh (1971) has presented empirical evidence that the probability of a company's being taken over, given its profitability, does decrease significantly with its size.[9] Now size should not be an impediment to a takeover in a perfect capital market. But the capital market cannot operate perfectly if there are asymmetries in the distribution of the relevant information, see Stigler (1967). Such asymmetries are precisely what we can expect managers to seek and create to protect their positions, and these would appear to correlate positively with size and growth.

Jensen and Meckling (1976) have developed a model of the managerial firm in which managers do have some discretion to pursue their own goals, but are induced to reveal information about their company's performance to raise capital. The predictions of their model would, thus, seem to be at odds with ours, and also, fortunately, with reality. The Securities and Exchange Commission was founded following the Great Crash of 1929, which revealed that many corporate managers had concealed information from investors, which furthered managerial interests at the expense of bond and stockholders. Since its inception the SEC has fought an on-and-off battle with corporations to induce their managers to reveal more information to which they are privy. The most recent round of this battle has been over the reporting by large diversified corporations of sales, profit and similar operating data by corporate division. This is precisely the kind of information one would not expect managers to reveal if they feared a takeover attempt, of course. Indeed, the reason why the company has diversified may be to conceal it. The situation in Europe is, if anything, worse.

The reason why corporations do not have to reveal information of this type to raise capital, as Jensen and Meckling predict, is that most corporations are not heavily dependent on the external capital market for investment funds. This is particularly true of large, mature companies. Thus, reliance on internal fund flows as a source of investment capital is complementary to a management's goals of preserving its power vis-á-vis the other factor owners.[10]

Once again, analogous arguments can be extended to regulated firms, nonprofit organizations, and government bureaucracies. William Niskanen (1971) develops his model of bureaucracy on the assumption that bureaucrats are self-interested individuals, who maximize the size of their budget. Niskanen gives little justification for the latter behavioral assumption. The theory presented here helps to explain why this is a plausible goal and in so doing links Niskanen's theory to the traditional literature on bureaucracy extending back to Weber. Increasing the size and complexity of a bureaucracy should increase the insider-bureaucrat's control over information relative to that of its monitors, thus increasing the bureaucrat's power to achieve his personal goals, whatever they might be.

## IV. Hierarchy, Power and the Distribution of Profits [11]

The traditional way of dealing with situations of uncertainty, in which one party may be able to take advantage of another, is for the parties to form a contract specifying the rights and obligations of each under the various contingencies that may arise as time unfolds and the uncertainties disappear. Should conflicts arise at some point, the parties to the contract can then appeal to an impartial third party to arbitrate their claims as established and guaranteed under the contract. Given the uncertainties and potential for conflict over the distribution of residual share that exist in the firm, one would naturally expect the members of this team to resort to the use of contract to protect their claims to the residual share. The major factor

owners of the firm are, of course, joined in a form of contractual relationship. Let us see therefore how information and uncertainty are handled under these contracts.

The contract between the common shareholders and the corporation is decidedly open-ended. Although the profits of the company figuratively belong to the stockholders, the determination of what gets reported as profits is made by the management, and the determination of what fraction of those profits that are reported gets paid as dividends is made by the board of directors. In principle, this latter body is supposed to serve as an impartial third party between management and stockholders to ensure that management does not abuse its insider's position at the expense of the stockholders. In practice it is typically under management control with management occupying several positions on the board, and undoubtedly wielding more power than its numbers suggest due to the greater amount of information its representatives possess about company operations. Indeed, since the board is heavily dependent on the management for information it must be largely under management's control. This fact is revealed in the following quote from the Board of Directors of the Pennsylvania Railroad made more than 100 years ago, and some 50 years before Berle and Means (1932). "The present form of organization (part-time directors and full-time officers) makes practical ciphers of the Directors, and this is from no deliberate intention, but from the very necessities of the case." After presenting this quote, Alfred Chandler (1962, p 313) went on to observe that, "Once a large business had reached a size that required the services of several full-time administrators, the board and the stockholders had only a negative or *veto* power on the government of their enterprise and on the allocation of its resources. They could say no, but they had neither the information nor the awareness of the company's situation to propose realistic alternative courses of action."

Thus, the stockholder's contract with corporate management does not offer much protection against the management's power to claim a larger fraction of profits than was understood at the time both became parties to the contract. The stockholder's major means of controlling management remains his right to sell his shares, or refuse to buy. We are thus back to the threat of takeover, and the discipline of the capital market. It should be stressed that either of these would suffice if the management could not withhold information

from the market. Solow (1971) has shown that the capital market can discipline a growth-maximizing management to maximize stockholder welfare by withholding capital from the company at the time of its inception. To do this, however, the market must know at the time the company is born what its growth and investment pattern will be throughout its entire life. Armed with this amount of information, the capital market has full control (power) over management. But obviously the uncertainties which surround a company's future at its birth allow no such discipline. As time passes, and the future becomes the present, information on how the various uncertainties facing the company are being resolved accrues asymmetrically to management and the stockholders. This unbalanced accumulation of information shifts the balance of power in favor of management and allows it to interpret the terms of the stockholder-management contract in a way which is most favorable to the latter.

As John Commons (1924, p 285) once observed, the wage contract typically "is not a contract, it is a continuing implied *renewal* of contracts at every minute and hour based on the continuance of ... satisfactory service ... and compensation'. The chief, and often only, explicit stipulation of the contract is that the employee agrees to accept authority within some limits for a certain wage. See Simon (1957). Thus, the labor contract is open-ended with respect to both time and duties. Given that labor and management participate in a teamwork activity, the fruits of their participation will appear as a joint product the division of which is in part arbitrary. The potential thus exists for conflict between worker and management over how the jointly produced residual is divided (wages), and how the vague limits to managerial authority are determined in practice.

The nature and complexity of the employment relationship requires that the bulk of the terms of this contract remain vague and implicit. Indeed, as Wiliamson (1975) has emphasized, implicit contracts are the distinguishing feature of hierarchical organizations. But with the bulk of the terms of the contract implicit, they cannot be arbitrated by impartial third parties. Instead, one of the parties to the contract must itself arbitrate the contract, and this task naturally falls to management given its role of information gatherer and monitor. But, this also gives management great latitude to interpret (arbitrate) the contract in ways most advantageous to itself.

The worker's ability to ensure his share of the

company's joint product comes not from his ability to enforce the terms of contract on an ongoing basis, but, in much the same way as with stockholders, in his right to quit or not join the company. Economists often assume this right suffices. Alchian and Demsetz (1972), for example, compare the worker-manager relationship to the customer-grocer relationship. But labor is seldom as mobile as this analogy suggests. Softness in the labor market, the accumulation of industry or firm specific skills, or merely the inertia resident in a given community builds up over time, produce rents that can be appropriated by management. The only way for labor to protect itself from this form of exploitation is to demand more explicit contracts amenable to third party arbitration.

It is interesting to note in this regard that employment contracts in the public sector have typically been much more specific and protective of employee rights than they have been in the private sector. Why this should be so is not clear. Public sector employees would not appear to be inherently less mobile, as a group, than private sector employees, although in some areas the government is in a monopsonist position, and conceivably could exploit employees with "firm" specific human capital. Whatever the explanation, it is interesting to observe the extension of civil service-type rules from the public to the private sector, a trend more pronounced in Europe than America but nevertheless observable there also. The growth of labor unions and the strengthening of the labor contract is also in part a method for increasing the worker's capacity to monitor managers, in part a formalization of the worker-management contract to allow third party arbitration. Finally, experiments in worker participation are efforts to involve workers directly in information gathering and the mutual monitoring of managers. Not surprisingly these have met with the greatest resistance from management.

## V. Conclusions

Since the end of World War II both governments and business have grown tremendously in size. What is more, in most countries this appears to be a continuation of a secular process rather than the outgrowth of a cycle. In the United States at least, the growth of the former has recently been treated with some alarm. Social scientists from a variety of disciplines and ideological persuasions have begun to explore models of budget-maximizing bureaucrats and vote-maximizing legislators. Somewhat surprisingly, the growth in business size has not met with a similar reaction. Economists, in particular, to the extent that they notice it at all, appear to treat it as the natural consequence of the Darwinian forces of the market seeking out more efficient organizational forms.

The arguments of this paper suggest that the two phenomena may be more closely related than generally believed. The differences between profit and nonprofit institutions may be far less than seems to be implied by the economics literature, at least. All organizations must deal with uncertainty, all must gather and process information to do so. In the process, certain individuals within the organizations will be vested with the power to advance their own goals to the disadvantage of other members of the organization, and can be expected to exercise that power. These characteristics all organizations have in common, and they can be expected to produce important similarities in their performances. Rather than continually stressing the difference between profit-oriented and nonprofit-oriented bureaucracies, we might begin now to explore some of their similarities.

## Footnotes                       •

*   University of Maryland, College Park.
1   The importance of secrecy to creating and preserving political power has been emphasized by Francis E Rourke (1961, 1969). See also Mills (1956) and Weber (1947).
2   See Amacher, Tollison and Willett (1976) and Rourke (1969, pp 55–58). The military is one of the three major groups making up Mills' (1956) power elite. Mills also lays great emphasis on the importance of secrecy in maintaining power.
3.   The most extensive development of the uncertainty-based theory of profit is by Knight (1921).
4.   Schumpter did not speak of information but of innovations. They amount to the same thing, however. For an innovation is nothing more than an idea that a new product (invention), or process, or organizational structure will produce a profit. And it is successful only to the extent that the idea (information) is a good one. Schumpeter's theory is best developed in A Theory of Economic Development (1934).
Other important theories of the firm and of profit can also be related to information of a specific kind. Thus, Ronald Coase stresses information about what kinds of activities are more efficiently handled within the firm than in the market (1937). For further discussion see Mueller (1976).

5. On the link between managerial salaries and profits, see Lewellen and Hutsman (1970) and Masson (1971). On insider trading see Manne (1966).

6. See, e.g. Downs (1967).

7. See, e.g. Baumol (1967), Marris (1964), Williamson (1964) and Galbraith (1967).

8. Several writers have sought an analogue for power in the economic sphere. Peter Blau (1964) compares the Knightian entrepreneur's receipt of profit to the political leader's receipt of power as reward for making risky decisions, but does not develop the analogy. Talcott Parsons (1963) compares power to money.

9. Kuehn (1975) and Smiley (1976) present additional evidence regarding the slack in the takeover mechanism. Smiley's results are particularly interesting. He found that a successfully takenover firm had fallen to 50 percent of its potential value by the time of its takeover, but that only 30 percent of this loss appeared to be recoverable following the takeover. Thus, as one might expect, the managers exercised their claim on the company's profits in such a way as to limit the gains from successful takeover to a fraction of their potential magnitude.

10. For further discussion of the relation between the managerial theory of the firm and internal investment theories see Grabowski and Mueller (1972). On the importance of firm maturity to the stockholder/manager conflict see Mueller (1972) and Grabowski and Mueller (1975).

11. This section draws in part on FitzRoy and Mueller (1979).

## Bibliography

Alchian, A A and Demsetz, H, 1972, "Production, Information Costs and Economic Organization", *American Economic Review*, 62, pp 777–95.

Amacher, R C, Tollison, R D and Willett, T D, 1976, "Risk Avoidance and Political Advertising", in their *The Economic Approach to Public Policy*, Ithaca, Cornell University Press, pp 405–33.

Bell, R, Edwards, D V and Wagner R H, eds, 1969, *Political Power*, New York, Free Press.

Berle, A A and Means, G C, 1932, *The Modern Corporation and Private Property*, New York, Commerce Clearing House.

Blau, P M, 1964, *Exchange and Power in Social Life*, New York, Wiley.

Chandler, A D, 1962, *Strategy and Structure*, Cambridge, Massachusetts, MIT Press.

Coase, R H, 1937, "The Nature of the Firm", *Economica* 4, pp 386–405.

Commons, J R, 1924, *Legal Foundations of Capitalism*, New York, Macmillian.

Crozier, M, 1964, *The Bureaucratic Phenomenon*, Chicago, University of Chicago Press.

Dahl, R A, 1957, "The Concept of Power", *Behavioral Science*, 2, pp 201–15, reprinted in Bell, Edwards and Wagner, 1969, pp 79–93.

Downs, A, 1957, *An Economic Theory of Democracy*, New York, Harper and Row.

Downs, A, 1967, *Inside Bureaucracy*, Boston, Little, Brown & Company.

FitzRoy, F R and Mueller, D C, 1978/79, "Contract and the Economics of Organization", *Siena Series*, Number 8, Acton Society.

Galbraith, J K, 1967, *The New Industrial State*, Boston, Houghton Mifflin.

Grabowski, H G and Mueller, D C, 1972, "Managerial and Stockholder Welfare Models of Firm Expenditures", *Review of Economics and Statistics*, 54, pp 9–24.

Grabowski, H G and Mueller, D C, 1975, "Life-Cycle Effects on Corporate Returns on Retentions", *Review of Economics and Statistics*, 57, pp 400–09.

Jensen, M C and Meckling, W, 1976, "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure", *Journal of Financial Economics*, 3, pp 305–60.

Knight, F H, 1965, *Risk Uncertainty and Profit*, New York, Harper and Row. (1st ed 1921).

Kuehn, D, 1975, *Takeovers and the Theory of the Firm*, London, Macmillan.

Lewellen, W G and Huntsman, B, 1970, "Managerial Pay and Corporate Performance", *American Economic Review*, 60, pp 710–20.

Manne, H G, 1966, *Insider Trading and the Stock Market*, New York, Free Press.

Marris, R, 1964, *The Economic Theory of Managerial Capitalism*, New York, Free Press.

Masson, R L, 1971, "Executive Motivations, Earnings and Consequent Equity Performance", *Journal of Political Economy*, 79, pp 1278–92.

Mills, C W, 1956, *The Power Elite*, New York, Oxford University Press.

Mueller, D C, 1972, "A Life Cycle Theory of the Firm", *Journal of Industrial Economics*, 20, pp 199–219.

Mueller, D C, 1976, "Information, Mobility and Profit", *Kyklos*, 29, pp 419–48.

Niskanen, W W, Jr, 1971, *Bureaucracy and Representative Government*, Chicago, Aldine-Atherton.

Parsons, T, 1963, "On the Concept of Political Power", *American Philosophical Society*, 107, pp 232–62, reprinted in Bell, Edwards and Wagner (1969, pp 251–84).

Rourke, F E, 1961, *Secrecy and Publicity*, Baltimore, John Hopkins Press.

Rourke, F E, 1969, *Bureaucracy, Politics and Public Policy*, Boston, Little, Brown & Company.

Russell, B, 1938, *Power*, New York, Norton.

Schumpeter, J A, 1934, *A Theory of Economic Development*, Cambridge, Massachusetts, Harvard University Press.

Simon, H A, 1953, "Notes on the Observation and Measurement of Power", *Journal of Politics*, 15, pp 500–16, reprinted in Bell, Edwards and Wagner (1969, pp 69–78).

Simon, H A, 1957, *Models of Man*, New York, Wiley.

Singh, A, 1971, *Take-overs*, Cambridge, Cambridge University Press.

Smiley, R, 1976, "Tender Offers, Transactions Costs and the Theory of the Firm", *Review of Economics and Statistics,* 58, pp 22–32.

Solow, R M, 1971, "Some Implications of Alternative Criteria for the Firm", in R Marris and A Woods, eds, *The Corporate Economy,* Cambridge, Massachusetts, Harvard University Press, pp 318–43.

Stigler, G J, 1967, "Imperfections in the Capital Market", *Journal of Political Economy,* 75, pp 287–92.

Wagner, R H, 1969, "The Concept of Power and the Study of Politics", in Bell, Edwards and Wagner, pp 3–12.

Weber, M, 1947, *The Theory of Social and Economic Organization,* ed T Parsons, New York, Free Press.

Williamson, O E, 1964, *The Economics and Discretionary Behavior,* Englewood Cliffs, New Jersey, Prentice Hall.

Wiliamson, O E, 1975, *Markets and Hierarchies,* New York, Free Press.

# The Economic Position of the Enterprise in the Yugoslav Economy

BY SVETOZAR PEJOVICH*

*The purpose of this paper is to explain the formation and distribution of total revenue in Yugoslav enterprises. The paper suggests the way in which the prevailing legal and institutional constraints, and internal rules affect the firm's output, employment and investment decisions. It concludes with a few remarks concerning built-in inefficiencies of the Yugoslav system of self-management.*

## Introduction

Economic and social institutions in Yugoslavia deserve our attention for several reasons. First, the development of the Yugoslav economic system covers a thirty year period in which, by trial and error, a unique set of institutional arrangements has been hammered out. Secondly, the analysis of the labor-managed firm in Yugoslavia provides some useful insights into the possible economic and social consequences of the codetermination movement in Western Europe.[1] Thirdly, the Yugoslav economic system is perhaps the only serious attempt in the socialist world to devise an effective incentive and control system to direct production without a large planning bureaucracy.

In Yugoslavia, decisions concerning the quantity and quality of output, employment, pricing, and the rate of capital formation are made by business firms themselves. It follows that to explain and understand the character of economic life in that country, a theory of the Yugoslav firm must be developed.[2]

The purpose of this paper is to discuss the formation and distribution of total revenue in Yugoslav enterprises. Analysis of total revenue points to the way in which the prevailing legal and institutional constraints, and internal rules of the game affect the firm's output, employment and investment decisions. In addition, the analysis of the firm's total revenue reveals some unique features of the Yugoslav economic system.

The paper begins with a brief description of the institutional framework in Yugoslavia that is relevant for discussion in the paper. The second section explains the firm's legal and institutional constraints, and identifies their behavioral consequences. The third section provides some general statistical evidence on the economic situation of business firms in Yugoslavia. The paper ends with a few concluding remarks.

## I. The Yugoslav System of Contractual Self-Management

The social and economic basis of self-management in Yugoslavia lies in (i) the state's right of ownership in capital goods, (ii) the employees' right to approve, police and enforce the decisions made by the managing organs in their respective institutions, and (iii) the system of contracts among self-managing organizations.

The term "firm" (*preduzece*) has been almost completely eliminated from the Yugoslav legal jargon. Instead, the Yugoslav literature uses the expression: organization of associated labor. The term *associated* labor refers to all economic activities that combine live labor with socially (state) owned assets. Legal substance of associated labor lies in the fact that the returns from capital goods can be captured only by those who combine their live labor with physical assets. The term organization of associated labor embraces all organizations that are (i) self-managed, and (ii) carry out their activities with socially owned capital goods. An organization of associated labor "... is in fact what was earlier referred to as a firm in the economic sector and an institution in the non-economic sector".[3]

The fundamental economic unit in Yugoslavia is the *basic organization of associated labor*. It is a

component part of an organization of associated labor. Basic organizations of associated labor were in the past referred to as work-units, plants and departments.[4]The terms organization of associated labor and basic organization of associated labor are very broad and cumbersome. For that reason the paper will use more conventional terms: firm and plant, but these terms are to be understood in the much broader content which they have in Yugoslavia. A firm might have one or more plants. Since the plant is the basic productive unit in Yugoslavia, the analysis.here is concerned with the plant's total revenue.

The law says that if the results of their joint labor can be measured in terms of value in either the market or within the firm and can be independently expressed, the employees should form their own plant (shipping and receiving in a manufacturing firm, sales personnel in a department store). The identifiability and separability of the flow of receipts is the major factor here. Moreover, the employees can take the plant out of the firm provided that the benefits they are expected to receive from leaving exceed the cost borne by other plants who remain with the firm.[5] In case the new plant leaves the firm, the latter is entitled to recoup only its investment expenditure adjusted for inflation.

The right to govern a plant is vested in all of its employees. The employees exercise their right of governing the plant in two ways: indirectly and directly. The forms of direct control are general meetings and referendums. Indirectly, the employees govern through the Workers' Council – the highest organ of the management in the basic organization. Members of the Workers' Council are elected by all the employees for a two-year term. Members of the Council continue at their regular jobs, they receive neither additional compensation, nor offices, nor privileges while serving on the Council. The director of the plant is appointed (and fired) by the Workers' Council. The position must be filled by way of a public contest. When the firm has more than one plant, each plant sends its representatives to the firm's Workers' Council which, in turn, elects the director of the firm.

Plants that belong to the same firm negotiate a contract. The contract must specify their mutual rights and obligations. In general, the contract specifies the composition of decision-making bodies in the firm, regulates commercial relations between the plants, provides guidelines for the distribution of income, assigns costs of law suits and other damages, and coordinates production and financial plans of plants that belong to the firm. Self-management contracts regulate issues of common interest to the workers in enterprises, the employees in local communities, and other self-managing organizations in related activities of a region. Those contracts specify the pooling of resources for joint undertakings, criteria for the distribution of income within the participating organizations, and other questions concerning their cooperation. Social contracts are negotiated by firms, trade unions, trade associations and socio-political organizations in a region. The purpose of social contracts is to settle the issues of common interest in each region and replace the regulative role of the state in the resolution of some key economic issues. Finally, the provision of many services (health, education, power production) is negotiated contractually between plants, other institutions and citizens' groups that use those services, and those who supply them. Buyers and sellers of those services form contractual associations (a different one for each service) covering a region.[6] Those contractual associations are called self-managing communities of interest.

In fact, contractual agreements encompass the entire social and economic life in Yugoslavia. In one sense, contractual agreements are not voluntary because they are mandated by law, and the basic constraints are frequently stipulated in advance. However, within these constraints, the terms of contract are negotiated among the participants. An immediate consequence of the Yugoslav system of contracts is to reduce the role of the state in regulating and controlling economic life. A dynamic, and perhaps the most important, consequence of the system of contractual self-management is that it generates incentives for the participants to seek greater freedom in negotiating the terms of contractual agreements.

## II. The Formation and Distribution of Total Revenue

The employees in a Yugoslav firm have one important property right: they own the plant's net-product (Dohodak) after taxes. This property right makes the analysis of the formation and distribution of total revenue central to better understanding of the relationship between incentive structures on the one hand, and the plant's output, employment and investment decisions on the other.

Our discussion of the formation and distribution of total revenue is based on a concrete example.[7] The plant in question is engaged in wholesale and retail trade. The firm to which this plant belongs is located in Belgrade. The accounting period considered here is January 1–June 30, 1978. For reasons of convenience all amounts are expressed in U.S. dollars. At the time of writing the value of $1 was about 18 dinars.

*Turnover Tax.* The turnover tax is paid by consumers purchasing final output and does not enter into the plant's income statement. Therefore the plant under consideration pays turnover tax only on the value of its retail sales. It is difficult to specify the turnover rate because it combines federal, republic and local rates, and the last two vary from one region of the country to another. An interesting feature of the Yugoslav tax system is that only customs and turnover taxes are paid into the federal budget. All other taxes are paid into the budgets of individual republics, counties, and self-managing communities of interest. In the case under consideration the total turnover tax bill was $1,438,495 or 18.3 percent of retail sales ($6,421,161).

*Total Revenue.* The Yugoslav law recognizes three major souces of total revenue: sales of goods and services, the value of transactions between plants in the same firm, and returns on external investment (time deposits, credits to other organizations, and joint projects). The plant's total revenue *includes* expenditure and other obligations incurred in the past and collected during the accounting period (and excludes expenditure incurred during the accounting period) if this expenditure is contained in the value of output that is yet to be sold or is sold but still to be paid for. The plant reported the following revenues for the accounting period:

| Total Revenue: | | $21,405,972 |
|---|---|---|
| Sales of goods | $15,946,089 | |
| Internal Transactions | 3,988,733 | |
| External investment | 646,150 | |
| Other revenues | 825,000 | |

*Production Expenses.* Production expenses are expenditure on goods, expendable supplies, contractual services, transportation costs, advertising, regular maintenance of fixed assets, travel, costs of energy and reproductive materials, and other business expenses. The plant incurred the following production expenses from January 1–June 30, 1978:

| Production Expenses | | $18,725,928 |
|---|---|---|
| Purchase costs of goods | $17,142,172 | |
| Business expenses | 1,527,680 | |
| Bad debts | 6,158 | |
| Inventory adjustments | 21,844 | |
| Court costs and penalties | 1,797 | |
| Other expenses | 26,277 | |

*Depreciation.* In general, the plant must maintain the book value of its assets by reinvesting depreciation allowances. When an asset is sold to another organization, the proceeds from sale must also be reinvested. It means that sale of existing assets generates net investment in Yugoslavia. If the sale price is less than the asset's book value, the difference has to be made up from the plant's earnings. These requirements affirm the state's ownership rights. However, the government cannot take capital goods away from enterprise. The collective is free to change the composition of its assets. It can substitute one type of asset (time deposits) for another (machines). Periodically, fixed assets are revalued to reflect changes in the price level. In our case, the plant had the obligatory depreciation of $91,666.

| Depreciation | $91,666 |
|---|---|

*Dohodak.* A conceptually correct translation of the term *dohodak* is value-added or net product. *Dohodak* is equal to the plant's total revenue less production expenses and depreciation.[8] The working collective has incentives to minimize production expenses. There exists a sharp positive relationship between the workers' present and future incomes on the one hand, and the size of *dohodak* on the other. The republic and local governments also have incentives to monitor expenses incurred by enterprises. Their tax revenues depend on the size of *dohodak*. Indeed, the government closely monitors the financial transactions of Yugoslav enterprises. The watchdog is the Office for Social Bookkeeping. Yugoslav enterprises can make payments to others neither directly nor through banks. *All* financial transactions must be cleared through the Office of Social Bookkeeping. In fact, it is the Office for Social Bookkeeping that instructs banks to make payments on behalf of enterprises. However, the costs of detecting, policing and eliminating wasteful activities are not negligible. Thus, the Office for Social Bookkeeping is primarily concerned with the legality of payments (e.g. that monies from one fund are not used to finance transactions of another fund) rather

than their business justifications. At the same time, the director of the plant, the management group, and other influential members of the collective (Chairman of the Workers' Council, the Party leadership in the plant) can increase their total compensation *via* the consumption of non-pecuniary goods such as frequent business trips, sponsorship of conferences, use of company cars, banquets for business associates, and the like. While incentives to minimize production expenses exist in Yugoslavia, the positive cost of monitoring the director and his associates suggests that some unnecessary expenditures can be assumed.

The plant under consideration reported the following *dohodak* for the accounting period:

| | |
|---|---|
| Total revenue | $ 21,405,972 |
| less | |
| Production Expenses | 18,725,928 |
| Depreciation | 91,666 |
| *Dohodak* | $ 2,588,378 |

According to the Yugoslav law the plant allocates its *dohodak* to: fixed legal obligations, legal obligations that depend on the size of *dohodak*, contractual obligations, and the residual that belongs to the collective.

*Fixed Legal Obligations from Dohodak.* The plant had to meet the following fixed obligations from its *dohodak:*

| | | |
|---|---|---|
| Fixed Legal Obligations | | $ 44,595 |
| Insurance for fixed assets, company cars, etc | $ 20,333 | |
| Insurance against injuries at work | 8,088 | |
| Supplementary disability insurance | 1,123 | |
| Land tax | 10,441 | |
| Entertainment & gift tax | 2,444 | |
| Solidarity Fund tax | 2,166 | |

Entertainment tax is 20 percent of the plant's expenditure on entertainment and gifts. Contribution to the Solidarity Fund is 10 percent of the plant's expenditure on entertainment and advertising. The Solidarity Fund is to be used to alleviate damages caused by floods, earthquakes and other disasters. Fixed legal obligations stood at 1.72 percent of the plant's *dohodak.*

*Legal Obligations that Depend on Dohodak.* A unique feature of the Yugoslav tax system is that it specifies purposes for which tax payments are

to be used. An advantage of this approach is that it reveals the burden of various activities that are financed through taxes.

To calculate the plant's taxable *dohodak* several adjustments are made in its actual *dohodak* of $ 2,588,378. First, the plant is allowed to subtract from its *dohodak* the amount equal to the employee's guaranteed income. A worker's guaranteed monthly income is 55 % of the last year's average personal income in the country where he works. Clearly, guaranteed monthly income varies from one locality to another as well as from one year to another. In our case, the minimum guaranteed monthly income was $ 140.34 or $ 842 for the accounting period. The plant had 330 workers which means that it could subtract $ 277,860 from its *dohodak.* The next subtraction is the plant's contribution to the firm in the amount of $ 130,555. This is a contractual payment that all plants make to their respective firms for administrative, legal, accounting, and other services. The amount paid by the plant as obligatory loans for the development of less-developed republics and regions is also tax exempt. Each republic contributes annually to this fund 1.97 percent of its gross national product. To meet this obligation individual republics require each plant to contribute a percentage of the value of its monetary and physical assets net of the amount of loans already made in preceeding years. Loans are repaid to the enterprise in 15 annual instalments and carry 4 percent interest. Since Yugoslavia has had double digit inflation for some years those loans are in effect taxes imposed on Yugoslav enterprises. The plant under consideration paid $ 522,571 to this fund for economic development. Obligatory loans are paid from *dohodak.* The plant calculated its taxable *dohodak* as follows:

| | | |
|---|---|---|
| Total Subtractions from *Dohodak* | | $ 1,322,836 |
| Guaranteed personal incomes | $ 277,860 | |
| Contribution to the firm | 130,555 | |
| Obligatory loans | 522,471 | |
| Interest payments and banking service | 309,701 | |
| Insurance for fixed assets | 20,333 | |
| Memberships of trade ass'ns | 10,869 | |
| Contribution to the building fund | 17,992 | |
| Meals for workers | 33,055 | |

Taxable *dohodak* was then:

$ 2,588,378 - $1,322,86 =

$ 1,265,542

The following taxes are paid from the plant's taxable *dohodak:*

| Total Taxes from *Dohodak* | | $ 310,181 |
|---|---|---|
| Republic's tax (7.0 %) | $ 89,473 | |
| Education tax (5.5 %) | 69,604 | |
| Tax for science (.69 %) | 8,732 | |
| Retirement and disability insurance (7.26 %) | 91,878 | |
| Tax for children protection (1.81 %) | 22,906 | |
| Health insurance (.68 %) | 8,605 | |
| Building fund tax (1.5 %) | 18,983 | |

With one exception, Republic's tax, all taxes listed above are paid into the budgets of self-managing communities of interest. The provision of many services such as welfare, health, education and retirement is negotiated contractually by organizations which represent those who supply specific services and organizations which represent those who demand them. They form self-managed organizations covering a region. For example, in the field of health, a self-management agreement on the formation of a self-managing community of interest is made between self-managing organizations and other institutions representing those who are eligible (by law) to receive health services on the one hand, and self-managing organizations which provide those services on the other (hospitals, clinics, medical institutes, pharmacies, etc).[9] Territorial boundaries of the regions are determined by such factors as the economic conditions of life in the area, homogeneity of its population, the area geography, and the availability of health services. The law specifies the minimum provision of services provided by self-managing communities of interest. However, contractual partners are free to negotiate the provision of additional services. They also negotiate the cost of services and determine taxes that would raise the required revenue. It means that taxes vary from one region to another as well as from one year to another. Importantly, education, health protection, retirement and some other public services in Yugoslavia are not provided through state budgets, and taxes for these services are not paid into state budgets.

Variable legal obligations amounted to 11.98 percent of the plant's *dohodak.*

*Contractual Obligations of the Plant.* Contractual obligations by plant include payments to the firm in support of joint administration (when the firm has only one plant this payment is not made), interest on credits, payments for banking services and the like. The plant under consideration made the following payments in this category of expenditures from *dohodak.*

| Contractual obligations | | $ 458,230 |
|---|---|---|
| Contribution to the firm | $ 130,555 | |
| Interest payment and banking services | 309,701 | |
| Membership of trade associations | 10,869 | |
| Contribution for Civil Defense | 5,308 | |
| Court costs and damages | 1,797 | |

Contractual obligations were 17.70 percent of the plant's *dohodak.*

*The Residual.* The residual is equal to the plant's *dohodak* less its legal and contractual obligations. In our case, the residual is:

$ 2,588,378 − ($ 44,595 + $ 301,181 + $ 458,230) = $ 1,775,372 or 68.59 percent of the plant's *dohodak.*

The residual belongs to the plant's employees. According to Yugoslav law, the residual must be allocated into the Wage Fund, the Collective Consumption Fund, the Reserve Fund, and the Business Fund. The Workers' Council determines the allocation of the residual between all the various funds as well as the distribution of the Wage Fund among the individual employees of the plant. However, the Workers' Council must publicly announce its distributional criteria well in advance or have them approved by the collective at a general meeting.

In deciding the scheme for the distribution of the plant's residual, the Workers' Council is expected to adhere to distributional guidelines stated in self-management agreements.[10] As was pointed out earlier, self-management agreements are contracts negotiated by self-managing organizations in related activities of a region. For example, the 1977 self-management agreement for *Vojvodina* (an autonomous province) provided the

following distributional guidelines for plants engaged in trade;[11]

Average contractual residual per adjusted worker
$2,567

Lowest contractual residual per adjusted worker (63 %)
1,617

Highest contractual residual per adjusted worker (214 %)
5,494

If the realized average residual per adjusted worker is between $ 1,167 and $ 5,494 the plant should use the following guideline for the allocation of its residual

$$\frac{\text{Business Fund and Reserve Fund}}{\text{Residual}} =$$

$$= 57.0 - \frac{3400}{q}, \text{ where}$$

$$q = \frac{\text{realized residual per adjusted worker x 100}}{\text{avg. contractual residual per adj. worker}}$$

If the realized average residual per adjusted worker exceeds $ 5,494 the guideline is

$$\frac{\text{Business Fund and Reserve Fund}}{\text{Residual}}$$

$$= 100 - (260 - \frac{21400}{q}) \frac{77}{q}$$

These guidelines suggest that the ratio of the wage fund and collective comsumption fund to the plant's residual falls as earnings per adjusted worker increase. The plant must allocate at least 4 percent of its residual to the business and reserve funds, regardless of the size of dohodak.

The term *adjusted* labor is a common denominator for measuring the size of the labor force in individual enterprises. The common denominator is needed in order to make adjustments for the anticipated effects of the differences in human capital on residual-labor ratios in various plants. The method used in Yugoslavia is to attach a coefficient to each different level of human capital. The adjusted labor force of the plant is obtained by multiplying the number of employees in each category by the relevant coefficient and adding up the results. The coefficients used in *Vojvodina* in 1977 were as follows:[12]

| | |
|---|---|
| unskilled worker | 1.00 |
| semi-skilled worker | 1.20 |
| less than high school (or equiv) | 1.20 |
| qualified workers | 1.70 |
| high school (or equiv) | 1.70 |
| highly qualified workers | 2.10 |
| junior college (or equiv) | 2.30 |
| college (or equiv) | 3.00 |
| masters degree (or equiv) | 3.30 |
| Ph D (or equiv) | 3.80 |

The plant in our example had 330 employees in the period January–June 1978. However, we have no information on the size of its adjusted labor force. We can only assume that the plant did quite well. Evidence is as follows: the plant's annualized average residual per *unadjusted* worker was $ 10,760 in 1978, and the Workers' Council allocated 43 percent of its residual to the business and reserve fund. According to the self-management agreement for Belgrade (covering all industries) the average residual per *adjusted* worker was $ 2,644 in 1977, and the allocation to the business and reserve funds (consistent with the average residual per adjusted worker) was 17 percent.

The allocation to the wage fund and the collective consumption fund enhances the employees' current incomes. The residual allocated to the business fund adds to their future income.[13] The allocation of the residual is an important decision that has future value consequences for the employees. They capture the benefits and bear the costs of that decision. Within the prevailing legal constraints, the employees are contractual partners in the team production process.

The allocation of the plant's residual for the accounting period was as follows:

| | | |
|---|---|---|
| Residual | | $ 1,775,372 |
| Wage Fund | $ 860,168 | |
| Collective Consumption Fund | 151,794 | |
| Business Fund | 698,701 | |
| Reserve Fund | 64,709 | |

*Wages.* The Workers' Council allocated $ 860,168 to the Wage Fund. A unique aspect of the economic system in Yugoslavia is that the plant employees are residual-takers. Yugoslav workers receive no contractual wage like their counterparts in the West and the East. Moreover, they receive no contractual wage plus a share of profit. The entire wage of a Yugoslav worker comes from and, in effect, depends on his plant's residual.

There is a minimum wage in Yugoslavia. It is equal to 55 % of the average wage in the area during the preceeding year. When the plant's residual is not sufficient to pay minimum wages, the enterprise can use its reserve fund. When the reserve fund is not sufficient to cover the minimum wage, the local community is supposed to help. Each month, workers receive *advances* which are based on either their earnings in the previous year or the plant's financial plan. At the end of the accounting period, workers receive the balance. When the balance is negative, the difference is covered from the plant's *dohodak* in the next accounting period.

The distribution of the Wage Fund among the plant's employees is regulated as follows: The Workers' Council attaches a certain number of points to each position in the plant. The criteria used to determine the number of points for each job usually include skill required, education, health risks, hardship, working conditions and the like. Those are general criteria used in Yugoslavia but their relative importance differs from one enterprise to another. In addition, self-management agreements provide general guidelines for individual Workers' Councils. Importantly, job classifications and the number of points attached to each job in a plant must be publicly announced well in advance. Plants belonging to the same firm are also expected to negotiate some redistribution of their respective residuals in order to alleviate large income inequalities. The plant's wage fund after taxes is divided by the total number of points. Next, the value of a point is multiplied by the number of points associated with each job, and the result is the employee's take home income for that accounting period.

The plant pays taxes from the Wage Fund *before* the employees are paid. Taxes from the Wage Fund are primarily paid into the budgets of local political units and self-managing communities of interest. Those taxes vary from one region to another and their uses are specified. As was pointed out earlier, the Yugoslav tax system relates the benefits of various public activities financed through taxes to their costs. The plant under consideration paid the following taxes from its Wage Fund.[14]

| Total Taxes from the Wage Fund | | $ 273,103 |
|---|---|---|
| County tax (1.41) | $ 12,128 | |
| Social protection taxes (.52) | 4,472 | |
| Education taxes (8.11) | 69,760 | |
| Tax for cultural activities (.77) | 6,624 | |
| Tax for physical education (.16) | 1,376 | |
| Tax for children's protection (.88) | 7,569 | |
| Tax for employment services (.24) | 2,065 | |
| Retirement and disability (5.97) | 51,352 | |
| Health protection (7.89) | 67,867 | |
| Tax for land improvements (.74) | 6,365 | |
| Additional tax for health protection (1.0) | 8,602 | |
| Additional tax for disability (.20) | 1,720 | |
| Other taxes (3.86) | 33,203 | |

Taxes reduced the Wage Fund from $ 860,168 to $ 587,065. The average net income per unadjusted worker was then $ 1,779 for the accounting period, or $ 296 per month.

*Collective Consumption Fund.* The Collective Consumption Fund is used to finance the provision of some specific benefits for employees. Most common benefits financed from this fund are apartments for workers, purchase of apartments in resort areas, subsidies for cultural and sport events, scholarships for workers and potential employees, and investment in improved working conditions.

*Reserve fund.* According to Yugoslav law, all self-managing organizations must form the reserve fund. The monies in this fund are to be used to cover business losses, to finance retraining and job search for laid-off workers, to pay the minimum wage when the residual of a plant is not adequate, and to provide depreciation funds when an asset becomes obsolete before the end of its legal life.

Self-managing organizations must allocate to the reserve fund at least 2.5 percent of their *dohodak*. When the reserve fund in a plant reaches 25 percent of the average dohodak for the last three years, the plant's obligation to contribute to the reserve fund ceases. In our case, the plant allocated $ 64,709 ($ 2,588,378 X .025) to its reserve fund.

*Business Fund.* The business fund consists of the book value of the firm's fixed assets, working capital, and external investments. External investments are joint undertakings with other self-managing organizations, time deposits, credits and others. The purpose of the Business Fund is to carry out the process of production, repay loans (only principal because interest is a contractual obligation that is paid *before* the residual is determined), and finance new investments. The allocations to the business fund together with the plant's depreciation allowances forms the base for investment activity carried out by the plant.

The plant under consideration allocated $ 698,701 to its business fund. From this amount $ 522,471 was for obligatory locus, and $ 176,230 for the plant's business fund.

## IV. Concluding Remarks

The analysis of the prevailing institutional structures within which the Yugoslav plant operates and internal rules for the formation and distribution of total revenue suggests that the Yugoslav system of contractual self-management has some built-in inefficiencies. Let us specify those inefficiencies that could be readily inferred from this paper.

(1) The working collective in Yugoslavia has the right of ownership in the plant's residual but no right of ownership in the stock of capital. When a worker leaves the firm he loses all his claims to the future returns. It means the collective has incentives to choose investment projects (regardless of whether those projects are financed internally or via bank credit) that maximize net cash flows of the plant over the employees' expected employment with *that* plant. The employees will prefer those projects whose present value of the returns on the truncated flows exceed the present value of the outlays. In general, incentive structures in Yugoslavia favor investment decisions that maximize the near term cash flows over those alternatives that would maximize the plant's present worth.

(2) In Yugoslavia, old employees have to share with new workers the return from past investments. Yugoslav law has tried to alleviate this problem by allowing the enterprises to pay a bonus to old workers for each year of service. Yet, a share of returns from past investment is captured by new workers. New workers might also have different time horizons and different perceptions concerning the disposition of the resid-

ual. Thus, the marginal product of labor increment must be substantial and possibly in excess of the average product of labor if the increase in the plant's employment is to be acceptable to the original group. A testable proposition is that business enterprises have a conservative bias in hiring new workers. In 1977, the rate of unemployment in Yugoslavia was about 13 percent.

(3) The employees of a Yugoslav plant cannot transfer (sell) their claims to the plant's residual. The non-transferability of the workers' right of ownership in the plant's earnings means that the system provides no room for specialization in risk-bearing across individuals with different degrees of risk aversion. The Yugoslav workers are forced in the aggregate to bear risks which are, in fact, insurable by diversification.[16]

1 See M Jensen and W Meckling, "On the Labor-managed Firm and the Codetermination Movement in the West", paper presented at the Interlaken Conference on *Analysis and Ideology,* June 1977.

2 See E Furubotn, "The Long-Run Analysis of the Labor-Managed Firm: An Alternative Interpretation", *American Economic Review,* 66, March 1976, pp 104–23; S Pejovich, "The Labor-Managed Firm and Bank Credit, in *Economic Analysis of the Soviet-Type System* (ed J Thornton), Cambridge: Cambridge University Press, 1976, pp 242–254.

3 *The Constitution of the Socialist Federal Republic of Yugoslavia,* Belgrade: Sluzbeni List, 1974, p 307.

4 *Ibid.*

5 *The Law of Associated Labor,* Belgrade 1976, pp 111–2.

6 S Pejovich, *Social Security System in Yugoslavia,* Washington, D C: American Enterprise Institute, 1979.

7 The income statement discussed in this paper was published in a highly specialized journal whose purpose is to instruct accountants and managers on the implementation of Yugoslav laws and regulations. See *Contemporary Practice,* Appendix No 703, 1978.

8 For detailed discussion see E Furubotn and S Pejovich, "The Formation and Distribution of Net-Product and the Behavior of the Yugoslav Firm", *Jahrbuch der Wirtschaft Osteuropas,* 3, 1972, pp 265–88.

9 S Pejovich, *Social Security System in Yugoslavia,* op cit.

10 These guidelines were introduced in the early 1970s in response to a continous decline in the allocation of the residual to the Business Fund. The author was recently told by the director of a Yugoslav enterprise that those guidelines are being abolished. However, the author has been unable to verify it.

11 See *Bulletin for Trade, Catering and Tourism,* No 10, 1977, pp 55–6.

12 *Ibid.*

13 Allocations to the business fund generate higher wages in the future.

14 Taxes paid by the plant on its wage fund were not enumerated in the published income statement. Tax payments are calculated from a number of legal acts and verified with the director of an enterprise in Belgrade.

15 S Pejovich, "The Banking System and the Investment Behavior of the Yugoslav Firm", in *Plan and Market* (ed M Bernstein), New Haven: Yale University Press, 1973.

16 M Jensen and W Meckling, *op cit.*

# Efficiency and Transferability of Ownership Rights – A Comparison Between the Capitalist Firm and the Socialist Firm

BY PER-OLOF BJUGGREN*

*The efficiency implications of restrictions on the transferability of the ownership rights of the firm are examined in a comparison between a capitalist firm and a socialist firm of the Yugoslav labour-managed type. It is shown that lack of transferability in the latter case has several effects: (a) self-financing of firm investments through retained earnings becomes unattractive, (b) risky investments are shunned, and (c) the working of the market for corporate control is hampered. As an empirical example of the efficiency implications of restrictions on transferability the producer cooperatives in Swedish forestry and forest-based industries are examined. These producer-cooperatives are very similar to the Yugoslav labour-managed firms with respect to the transferability of ownership rights.*

## 1. Introduction

The literature on property rights holds that efficiency in the use of resources is promoted if such rights are well-defined, exclusive and transferable. The concepts of efficiency and property rights refer to the allocation of resources to their most valuable uses as determined by the preferences of ultimate consumers and the rights to use the resources in certain stipulated ways, respectively. (See e.g., Alchian and Demsetz [1937], pp 17–19, Cheung [1970], p 64 and Posner [1972], pp 10–12.) Exclusivity and transferability are the main characteristics of private property rights. (See Cheung [1978], p 51, Alchian and Allen [1974], p 142 and Pejovich [1971], p 144.) The owner of a private property right has the rights to exclude others from its use, to appropriate the income emanating from its use and to sell it on whatever terms he and the buyer find agreeable. (See Cheung [1978], p 51.) It is principally with respect to transferability that a public property right can be distinguished from a private property right. An inherent characteristic of the former is that it cannot be sold on a market. (See Alchian (1965), p 138.)

In a socialist state the means of production are publicly owned. As soon as the ownership rights to capital can be exchanged against other rights on the market, we have a regime of private ownership of capital – and the capitalistic state. This is so because transferability of property rights implies at least some degree of exclusivity in ownership (Cheung [1978], p 52) and, therefore, the fulfilment of both conditions stated above for a property right to be classified as private. With respect to the exclusivity of a public property nothing definite can be said. The individual citizen may or may not be permitted to use an asset exclusively and/or appropriate its yield. Examples are the Yugoslav firm and the Soviet firm. In the former, members are permitted to use and appropriate the yield from the use of capital goods exclusively, while in the latter they can use the capital good but cannot appropriate the yield. In both cases the quality of capital goods is assumed to remain unchanged, which means that the capital stock must be kept intact in perpetuity (Pejovich [1971], pp 143, 148, 152).

To highlight the significance of transferability as the distinguishing feature, a comparison is made here between the Yugoslav type of socialist firm and the corporation, taken as representative of the capitalist firm. The employee in the Yugoslav labour-managed firm and the shareholder in a corporation are on equal terms in the senses that they both have the ultimate authority in deciding on the uses of the firm's resources and they both have the right to the residual income, i.e. that remaining after all other parties to the

firm have received their contractual income. But while the Yugoslav worker cannot transfer his right, equity shares in a corporation can be bought and sold freely on the stock-market at whatever prices the seller and the buyer agree upon. The lack of transferability means, first, that the Yugoslav worker must stay with the firm to have a claim on the firm's income-stream whereas the shareholder in a corporation can obtain a price for his share which reflects the market valuation of the present value of that part of the future profit-stream to which the share gives a right. Furthermore, the Yugoslav worker cannot free himself from participating in management and sharing in profits/losses and become instead an ordinary wage-earner with a contractually agreed income as in a capitalist firm. Thirdly, there is no absentee ownership in the Yugoslav firm, since a person must work in the firm to share in management and profits/losses. Finally, there is a limit to the portion of the profit-stream that can be appropriated by one firm member, unlike the corporation where one individual can be a majority shareholder.

The lack of transferability will influence the efficiency of the economy in three different ways. First, self-financing of investments will be affected by the inability to capitalize the future profit-stream. Secondly, the supply of risk capital will be curtailed as there are no possibilities open to the individual to avoid risk by utilizing portfolio effects, by entering into employment contracts with a guaranteed wage or by in other ways diminishing his part of the responsibility for the conduct of the affairs of the firm. Thirdly, the working of the market for firm control will be hampered both by the lack of information about management efficiency as is supplied in a stock-market and by the fact that an outsider seeking control over the firm cannot capture more than a given part of the increment in the profit-stream which might result from his efficient management.

## 2. Self-Financing of Investments

The first aspect of efficiency to be considered is the self-financing of investments through retained earnings. The distortions that arise here can all be attributed to the inability to capitalize future profits through the sale of ownership rights. If we apply the efficiency rule that funds are to be retained for investment purposes when the rate of return obtainable on these funds promises to

be higher than that obtainable elsewhere in the economy, the capitalist firm, represented by the corporation, shows a propensity towards an excessive degree of self-financing. There are two reasons why this occurs. The first is simply that there may be non-pecuniary effects which, according to the preferences of the owners of the firm, outweigh the pecuniary loss measured by foregoing higher yielding investment opportunities elsewhere. Especially if the owners are active firm-members, non-pecuniary rewards in the form of consumption on-the-job will constitute strong arguments in their utility functions. Secondly, a lack of control due to dispersed and absentee ownership may mean that employed management can use the firm's earnings for discretionary expenses, which may have an investment character. This inclination on the part of employed management will be reinforced if the lack of control means that the managers are not rewarded according to their marginal productivity. (See Alchian and Demsetz [1972].)

Turning to the first of these two reasons it is clear that the excessive degree of self-financing is not a case of inefficiency. The property right holder, i.e. the owner of the firm, is simply using his right in a way that maximizes his utility. In the second case also it may be wrong to talk of inefficiency. Transaction costs in the form of costs to owners of policing and enforcing their exclusive rights to profits may be higher than the incremental yield per share that might result. (See Demsetz [1967], pp 357-9.) This problem will be compensated for and will, to a large extent, be overcome in a corporation by the high degree of transferability and the limited liability characterizing a share. The fact that corporations with dispersed ownership have hitherto met the test of the market and survived suggests that they represent an efficient type of ownership.

A socialist firm of the Yugoslav type represents the opposite case where the above efficiency rule would result in too low a level of self-financing. The absence of the possibility of capitalizing future profits through the sale of ownership rights and the obligation to keep the capital stock of the firm intact in perpetuity create an incentive for the Yugoslav worker to realise profits rather than reinvest them in the firm. Inasmuch as the workers really have the power to decide for themselves through the Workers Council the rate at which profits are to be realised in the form of higher wages, this is an incentive that will be catered to. (See Fu-

rubotn & Pejovich [1970], pp 32–33 and Furubotn [1971], p 183.) Besides, the scope for discretionary investment by management is circumscribed since workers can effectively control the management by making on-the-job observations of the way the firm is managed.

It follows that the shorter the time-period workers, on average, expect to stay with the firm, the higher must the rate of return on investments financed by retained earnings be relative to the returns which can be obtained elsewhere by the individual worker. In the Yugoslav economy, where no absentee ownership of the firm is allowed, the alternative open to the individual worker is essentially to put his money in a savings account.[1] If the average expected period of employment in the firm is 10 years and the rate of interest available on a savings account is 10 per cent and given that the capital stock is to remain intact for ever, a self-financed investment must give a return of at least 16 per cent in order to have a chance of being accepted by a majority of the workers. (See Pejovich [1970], pp 150–1.) (16 per cent corresponds to the annuity that has to be paid as instalments of a debt of 1 dollar running for 10 years at an interest rate of 10 %.) The shorter is the time horizon of a majority of the workers in the firm, the higher must be the required rate of return on investments financed through retained earnings compared to the rate of interest on a savings account. If we also take into consideration the possibilities of obtaining external funds, such as bank credits, at an interest rate lower than the required rate of return on self-financed investments, there will be no incentive to finance any investments through retained earnings. (See Furubotn & Pejovich [1973], pp 278–83.)

Inefficiency in the Yugoslav system appears in the form of an excessive degree of consumption. The inability to capitalize the profits from internally financed investments and the availibility of external funds for investments, obtainable at a relatively low cost, combine to encourage the realisation of profits. To a large extent the interest on a savings account is the yardstick against which the desirability of future relative to present consumption is judged, although there do exist higher yielding investment opportunities in the economy. It may be added that a high level of private consumption and a high demand for bank credits for investment purposes tend to create inflationary pressures which, in turn, make future consumption even more disadvantagous.

Another efficiency aspect to be considered in the context of self-financing is that the labour market tends to be rigid to an extent that is inconsistent with the allocation of labour to its most valuable uses i.e. where marginal productivity is at its highest. (See Furubotn & Pejovich [1973], pp 284–85.) If a worker has participated in the financing of investments through retained earnings he will take the loss of future yields from these investments into account when contemplating a change of occupation. The wage offered by another firm must be high enough to cover both the value of his marginal productivity in his present occupation and the income stream in the form of a share in future profits that he will loose. At the same time workers who have been sacrificing income for the financing of investments will be reluctant to let new workers enter and share in the yields from these investments. Marginal productivity considerations are ignored as a result of both the inability to capitalize future profits through sale of ownership rights and the absence of employment contracts with a guaranteed wage.

Justification for the rule that the firm's capital stock must be held intact may also be found in the inability to capitalize future profits. Without this qualification on the exclusive use of the firm's capital by the workers it could be expected that the workers would be motivated to consume the capital during their period of employment. Annual income might be raised during a shorter period of time through depletion of the capital stock, for example, by not undertaking replacement investment and by refusing to hire new workers, who might have a longer time horizon and who will constitute further claims on income.

## 3. Risk Bearing Effects

The risk to be considered in this section is primarily that due to demand and supply fluctuations which are outside the control of the firm i.e. risk factors of an exogenous character. The possibilities offered by a high degree of transferability of ownership rights for avoiding risk associated with inefficient management will be dealt with in the following section. Two types of transferability restrictions are of interest here, namely the lack of an opportunity for the worker to enter into an employment contract with a guaranteed wage and the prohibition of absentee ownership. These two transferability restrictions will result in risky business activities being shunned for three different reasons. First, the cost of acquiring risk

capital will be raised. Secondly, there will be no-one guaranteeing the wages of the workers. Thirdly, the worker cannot take advantage of portfolio effects in investments and cannot cater to differences in his preferences with regard to the types of business activities that he prefers because of the type of work offered and the business activities that he prefers for reasons of risk.

Few persons would be willing to lend at a fairly low fixed interest rate to a business promising high profits if successful and heavy losses if unsuccessful, where the probability of failure is high. The rate of interest charged on loans to such business activities will instead be high to compensate for the risk that the loans will never be repaid. (See Posner [1972], pp 1976–178.) To secure investment funds on more favourable terms it will be necessary to let the persons providing the funds share in the profits generated from successful operation and it may also be necessary to allow them some management control. This is exactly how the corporate firm proceeds when acquiring capital for the financing of risky activities. In a labour-managed firm of the Yugoslav type the right to share in the residual and the control of management are reserved for the workers of the firm. The only way open to the workers of a labour-managed firm to finance investments, besides providing the funds themselves from their own savings or from the profits of the firm, is to raise debts on the capital market (in principle, this is equivalent to bank loans).

This, due to the prohibition of absentee ownership, places the labour-managed firm in a dilemma. The workers may not be able to raise the necessary capital themselves or may be reluctant to place a major part of their savings in a risky business activity. Added to this is the impossibility of capitalizing future profits. But reliance on debt financing of investments will sooner or later increase the fixed costs of interest and amortization dramatically. Bankers will claim compensation in the form of a higher rate of interest and faster amortization, the greater the risk. The banks in Yugoslavia are also labour-managed firms and as such have an interest in not loosing money on the loans they are issuing. Risky business will tend to be shunned and where banks do become involved in financing risky business they will also demand some control over firm management.

The second reason is that the wages of the worker cannot be guaranteed. It is then especially disadvantageous for a worker to be employed in a risky capital-intensive industry. The restrictions

on transferability of interest in this context are the prohibition of absentee ownership and the lack of an opportunity on the part of the workers to enter into employment contracts with guaranteed wages. Since the worker cannot give up or sell his right to participate in the control of the firm and to share in the residual, it follows that he cannot escape the responsibilities associated with ownership of the firm. He must accept the obligation to bear losses as well as the opportunity to share in profits. Moreover, the effects of failure on the worker's income cannot be cushioned by having the losses spread over a larger number of absentee owners and by having special agreements on how absentee owners absorb losses (or share in profits). An example of a special agreement of this kind can be found in the use of ordinary shares and preferred ordinary shares in a corporation. Another comparison can be made with respect to the degree of responsibility for losses. In the capitalist corporation the shareholder has limited liability in that he is only responsible for losses up to an amount equal to the value of his shareholding.

In the Yugoslav system the liability of the workers is limited only through a bankruptcy law that declares a firm bankrupt if it cannot pay a certain minimum wage after the other contracting parties including the State, the banks and other suppliers of inputs to the firm have received their contractual share. (See Ward [1958], pp 568–569.) When this point is reached, the workers will find themselves unemployed unless the creditors agree to having the debts reduced. Therafter, the firm can be restored. It is clear that the risk facing the worker is considerable and inescapable. This must have an impact on the type of business investment preferred by labour-managed firms. Ideally, a labour-managed firm should operate under conditions of stable and easily forecast demand and supply. Given stochastic demand and supply conditions, there should preferably be a positive covariance between demand and supply. To the extent that these conditions are not fulfilled the worker's income will be subject to more or less severe fluctuations. Assuming he is risk-averse, the worker will wish to avoid this situation. To maintain stable consumption over time, he will be forced under such circumstances to keep reserves in the form of, for example, extra capital in a savings account. The worker will be particularly susceptible to uncoordinated fluctuations in demand and supply if the industry in which he is employed is capital-intensive.

The third reason is that individuals cannot separate risk-bearing from employment and, thereby, take advantage of portfolio effects in investments. The tying of risk-bearing to employment is a result of the restrictions on transferability imposed by having no absentee ownership and by the inability to enter into an employment contract with a guaranteed wage. The only risky investment a worker in Yugoslavia is permitted to undertake, apart from investment in human capital, private investments of the kind mentioned below and the choice of occupation, is to let his part of earnings be retained in the firm where he is employed, an investment opportunity which for reasons already discussed is rather unfavourable to the worker. There is no opportunity open to the worker to realise his part of the profits and reinvest the money in industries whose risk patterns he prefers either because he considers the risk of failure to be low or because there is a negative covariance between fluctuations in profit levels. Instead, the opportunity set for risky investments is, in practice, restricted to one point corresponding to the firm in which the person in question is employed. This lack of any opportunity to separate risk-bearing from employment makes it impossible for a worker employed in a risky line of business to take advantage of portfolio effects in ownership and diversify away the risk associated with the enterprise in which he is employed. This contrasts with the capitalist economy where there are no restrictions on absentee ownership of the corporation and where shares in a large number of different types of companies can be bought freely on the stock market.

## 4. The Market for Firm Control

The last aspect of transferability to be discussed is the possibility of obtaining ownership of a major portion of the firm through purchase of ownership rights. The type of efficiency relevant here is the extent to which control of productive resources rests with those persons best able to use resources in the most efficient ways. It has been argued that control of the firm can be considered as a valuable asset (Manne [1965]). By analogy with other assets, efficiency is equated with the most valuable allocation of control. An index of the degree of efficiency is the amount of profit that controlling individuals are able to produce for the firm. Profits may either be absorbed by the controlling management in the form of discretionary expenses or be distributed among the owners of

the firm, who may indeed be the top management of the firm. In a corporate economy changes in control over the firm can occur in three ways; through proxy rights, direct purchase of shares and mergers. The first and the third methods have counterparts in an economy with labour-managed firms. It is the second method i.e. outright purchase of shares in the stock markets, which distinguishes between the two economic systems.

In a corporate economy the stock market serves as a mechanism providing signals about management efficiency. A decrease in management efficiency tends to be reflected in a fall in the prices of the company's shares. The more the prices of the shares fall, the easier and cheaper it is to take over control of the firm. The stock market also acts in other ways as a vehicle for the dissemination of information about managerial efficiency. There are journals and brokers specializing in stock market affairs and as far as there is competition between stock exchange institutions, the survival of a stock exchange will be dependent on how well it succeeds in providing information about managerial efficiency.

The absence of a counterpart to the stock market as a provider of information does not, however, necessarily imply severe inefficiency. Before a signal is transmitted in the form of a fall in the price of the company's stock, information on managerial inefficiency is often already available to other firms. (See Manne [1965], p 119.) As a result of their more or less daily contact, customers and suppliers can form judgements about managerial efficiency and competitors can draw inferences from their knowledge of industry characteristics. In their roles as suppliers, customers and competitors, labour-managed firms in a socialist economy can obtain information about managerial efficiency in the same ways. The workers in a firm can make inferences about managerial efficiency by comparing their income with the income of workers employed by other labour managed firms in the industry. Problems will arise, however, if one firm considers that another firm could be run more efficiently but the majority of workers in the second firm cannot be persuaded about the desirability of a merger on the grounds of efficiency. We have in this case a problem of signalling parallel to that in the labour market when workers try to find a job with wage conditions corresponding to their perceptions of their marginal productivities. In some cases the cost if signalling will simply be too high to permit

the workers to obtain suitable employment. (See Spence [1974].) In the same way the cost of signalling might be too high for a potentially profitable merger to take place. In that case the transferability of ownership rights makes it possible for a corporation to take over control. The interested corporation simply buys a controlling part of the other company's stock. In the socialist economy high information costs of the type described here constitute an insurmountable barrier as outright purchase of ownership rights is forbidden.

If we consider the negotiation of a merger in the socialist economy it can be concluded that the lack of transferability again creates problems. Problems arise because there is nothing to negotiate about. The post-merger residual [2] in the combined firm is shared on a pro rata basis. This contrasts with the situation in the capitalist economy where the corporations themselves can decide on the rate at which the shares of the acquiring and the acquired firm are to be exchanged. In a socialist economy of the Yugoslav type the condition for a merger to take place is that the residual per worker must be higher in both the acquiring and the acquired firm, since no worker will vote for a merger that lowers his income. This conditions implies that the two following inequalities must be fulfilled for a merger to take place:

ners, i.e. firms being suppliers or customers, have widely different profit levels per worker these firms can be excluded as merging partners despite their informationally advantageous position.

Another aspect to be considered is that the incentive to undertake a merger is also dependent on how much of the increase in efficiency the new controlling management can appropriate. There are, in principle, three different ways through which the increment in efficiency can be appropriated, i.e. through a higher salary, through an increased opportunity set for discretionary spending and by capturing a larger or smaller part of the increment in the profit-stream of the merged firm through outright purchase of ownership rights. The first two ways will probably not differ to any great extent between the two property rights systems. In as much as information about managerial efficiency is spread throughout the economy, efficient managers can be compensated through higher salaries. Appropriation through discretionary expenses is also dependent on how well information about managerial efficiency is disseminated in the economy, but the influence operates in the opposite direction since the size of the opportunity set is dependent on the lack of outside control of managerial efficiency.

The third way of benefiting from increased efficiency is only open to managers in a corporate

$$\frac{y_2^1 + y_1^1}{n_1 + n_2} > \frac{y_1^0}{n_1}$$

$$\frac{y_1^1 + y_2^1}{n_1 + n_2} > \frac{y_2^0}{n_2}.$$

where:

$y_1^0$ = residual in the acquiring firm. Pre-merger situation

$y_1^1$ = "      "  "  "      "  . Post-merger situation

$y_2^0$ = "      "  "  "  acquired  "  . Pre-merger situation

$y_2^1$ = "      "  "  "  "      "  . Post-merger situation

$n_1$ = number of workers in the acquiring firm

$n_2$ = "      "  "  "      "  "  acquired  "

It follows that the difference in residual per worker between two firms must not be greater than the increase in residual per worker resulting from the increase in efficiency after the merger. We can then predict that mergers will be most likely to take place between firms within the same industry or between firms in industries that do not differ too much with respect to residual per worker. To the extent that vertically-related transactions part-

system. By purchasing shares the controlling management can directly appropriate a larger portion of the increment in the profit-stream attributable to increased efficiency. By having this option managers are not reduced to realizing the benefits of increased efficiency in the form of discretionary expenses if the labour markets fails to reward them appropriately through higher salaries. Thus, the incentive to take over control for efficiency

reasons will be enhanced in a corporate economy. If the first two means of appropriating the value of control cannot be relied upon, an individual always has the choice of increasing his share in the ownership of a firm and, thereby, appropriating the profits from an increase in efficiency. This is not possible in a labour-managed system where the purchase of ownership rights is prohibited.

## 5. The Effects of Restrictions on the Transferability of Ownership Rights – the Case of the Swedish Forestry and Forestry-based Industries

As an empirical illustration of the effects of restrictions on the transferability of ownership rights we have chosen the Swedish forestry and forest-based industries. Within this sector of the Swedish economy there are two types of firms (a) corporations and (b) producer cooperatives. The producer cooperatives exhibit great similarity to the labour-managed firms of the Yugoslav type with respect to the transferability of ownership rights. By making comparisons within the same industry and country it is possible to isolate the effects of restrictions on transferability in a way that is not possible when comparing a socialist and a capitalist state with their differences in economic and institutional conditions.

The Swedish producer cooperatives in the forest industry can be regarded as an analogue to the labour-managed firm where the firm's suppliers instead of the workers have the right to the residual and exercise ultimate control. (See Ståhl [1979], p 43.) The share of the residual that a supplier in the producer cooperative can appropriate stands in direct proportion to how much wood he has delivered relative to other owner-suppliers.[3] In principle, control consists of one vote per member-supplier in the election of the board of directors. By law (see Moberg [1966]) no-one other than owners of forests can obtain ownership rights in a producer cooperative in the forest industry. Furthermore, ownership is open to any supplier of wood in the region who wishes to join the producer-cooperative. Finally, unless otherwise decided by a majority of the owners, a supplier-owner has no right to receive more than the amount of the membership fee in the producer-cooperative when terminating his membership.

Drawing a comparison with the Yugoslav firm we find, first, that the supplier-owner has practically no possibility of capitilizing future profits of the cooperative through sale of his ownership

right. The prospective buyer of an ownership right must be the owner of a forest within the same region and as such he will have the option of obtaining the ownership right at no higher cost than the nominal membership fee. Unlike the Yugoslav worker, however, the supplier-owner has the right to pass his ownership right on to future generations or to his friends, provided that they are also owners of forests. In contrast to the worker-owner in the Yugoslav economy, the supplier of wood in the Swedish economy can enter into a guaranteed contract of delivery with a corporation. As in the Yugoslav firm, there is no absentee ownership since shares in the residual depend on the amount of wood delivered and no-one other than owners of forest can be members of the cooperative. Finally, there will also be in practice a limit to the portion of the residual that can be appropriated by any one supplier, the limit being set by the number of suppliers all of whom in the region are eligible for membership.

Of the different consequences of lack of transferability discussed in the paper it is only those of inflation, inefficiency in the labour market and a lack of portfolio opportunities that are not relevant here. Table 3 shows that the producer-cooperatives have relied primarily on capital from external sources to finance investment. Furthermore, it can be seen that the funds provided to the cooperatives have not been used for long-term investment purposes to the same extent as in the corporations. With no absentee ownership[4] the producer-cooperatives are confined to debt-financing of investments, unless the supplier-owners are willing to supply the necessary funds from their own pockets or the profit-level is high enough to permit financing through retained earnings. From Table 2 it can be seen that the cooperative firms have not been successful in attracting investment funds from their supplier-owners. Besides the inability to capitalize future profits, the lack of own capital (the low solidity) can be attributed also to the fact that forestry and forest-based industries represent a risky line of business. The degree of export orientation makes the business very sensitive to changes in trade conditions. Moreover capital intensity has been increasing steadily. (See e.g. SIND 1976:1, pp 140–170.) The reliance on external financing has meant that the producer-cooperatives have been very adversely affected by the rapid decrease in demand from its peak in 1974. (See Table 1.) The increase in fixed costs consequent upon debt financing has

shown itself in heavy losses during 1977 and 1978. Losses of the same magnitude cannot be found in the corporations operating in the same industries.

## 6. Summary

The main feature distinguishing a socialist state from a capitalist state is the absence of transferability of the ownership rights of the firm. The trademark of socialism is the public ownership of the means of production. The purpose of this paper has been to examine the efficiency implications of having no transferability of ownership rights. A comparison has been made between the socialist firm, represented by the Yugoslav labour-managed firm, and the capitalist corporation. While the shares of a corporation can be bought and sold freely on a stock-market, the ownership rights of a Yugoslav labour-managed firm are tied to employment.

The lack of transferability in the Yugoslav system shows itself in several ways: (a) the future profits of the firm cannot be capitalized, (b) the worker cannot free himself from the responsibility of bearing losses as well as sharing in profits (i.e. he cannot enter into an employment contract with a guaranteed wage), (c) no-one other than the workers has the right to share in the residual and exercise ultimate control over the management of the firm (i.e. there is no absentee ownership), and (d) the firm's residual is divided on a pro rata basis among the workers.

The inability to capitalize future profits has a negative effect on the degree of self-financing of firm investment and makes the labour-market rigid. The low degree of self-financing tends to give rises to inflationary pressures in the economy. The lack of an opportunity for the worker to enter into an employment contract with a guaranteed wage and the prohibition of absentee ownership result in risky investments being shunned. There is no way in which the consequences for the worker of adverse business conditions can be cushioned. A worker is especially susceptible to changes in business conditions if he is employed in a capital intensive industry. Finally, we find that the working of the market for firm control is hampered because of the lack of an opportunity to appropriate the profits from more efficient management by acquiring the right to a major portion of the firm's profit-stream.

The Swedish forestry and forest-based industries provide empirical illustration of the theoretical propositions of the paper. In the Swedish forest industry we can find two types of firms, corporations and producer-cooperatives. The producer-cooperatives show great similarities to the Yugoslav labour-managed firms with respect to the transferability of ownership rights. The principle difference is that private ownership of means of production is not forbidden in the Swedish economy and, therefore, a forestry-producer can always enter into a guaranteed delivery contract with a corporation. Empirical evidence for the producer-cooperatives is consistent with our arguments that (a) in comparison with the corporations financing through debts is prevalent, (b) the degree of self-financing of investments through retained earnings is lower and (c) the residual fluctuates widely in response to changes in business conditions.

## Footnotes

* University of Lund. I am indebted to Ingemar Ståhl for valuable comments. I am greatly indebted to James Love for editing advice on the wording of this article.

[1] According to Furubotn and Pejovich (1973, p 279) the investment alternatives open to the Yugoslav workers "are restricted to monetary assets, human capital and some very limited types of physical assets such as small shops, restaurants, taxi businesses, jewelry etc, where the right of ownership does not necessarily and obviously violate the principle of exclusive public ownership of capital goods."

[2] The term residual is used to take into account the fact that in a socialist economy different wage levels for different industries can be considered as normal because of differences in factors such as investment in human capital and working conditions.

[3] The residual will be reflected in the prices for wood that the cooperative is charging its members and in a possible refund in proportion to the quantity of wood delivered if business turn out to be better than expected.

[4] It is, however, to be noted that the problem of no absentee ownership in producer-cooperatives has to a large extent been solved by having affiliated companies that are run in a corporate form (see SOU 1979:11, pp 28–29 and 34).

Appendix: *Financial Data for the Swedish Forest Industry. The tables cover the 18 largest forest companies.*

*Table 1. Profitability*

Profitability 1973–1977, all 18 companies, million *kronor*

|  | 1973 | 1974 | 1975 | 1976 | 1977 |
|---|---|---|---|---|---|
| Sales | 17.500 | 25.100 | 23.900 | 26.300 | 28.300 |
| Gross profit = profit before deprecia- tion | 3.400 | 7.000 | 4.300 | 3.300 | 1.500 |
| Net profit = profit after depreciation and financial items | 1.800 | 5.400 | 2.500 | 1.00 | −1.600 |
| Gross profit (%)[1] | 19,4 | 27,9 | 18,0 | 12,5 | 5,3 |
| Net profit (%)[1] | 10,3 | 21,5 | 10,5 | 3,8 | −5,7 |

Profitability 1973–1978, producer cooperatives, million kronor

|  | 1973 | 1974 | 1975 | 1976 | 1977 | 1978 |
|---|---|---|---|---|---|---|
| Sales | 3.116 | 3.890 | 3.674 | 4.407 | 4.976 | 5.394 |
| Gross profit | 469 | 898 | 567 | 455 | −165 | −89 |
| Net profit | 191 | 624 | 254 | 17 | −755 | −799 |
| Gross profit (%)[1] | 15,1 | 23,1 | 15,4 | 10,3 | −3,3 | −1,6 |
| Net profit (%)[1] | 6,1 | 16,0 | 6,9 | 0,4 | −15,2 | −14,8 |

[1] In percentage of sales

Source: Regeringens proposition 1978/79:207.

*Table 2. Financial structure*

Financial structure, all 18 companies

|  | Million kronor | | | In percentage terms | | |
|---|---|---|---|---|---|---|
|  | 1972 | 1974 | 1977 | 1972 | 1974 | 1977 |
| Short term debts | 4.500 | 6.800 | 10.400 | 20 | 22 | 22 |
| Long term debts | 9.100 | 9.700 | 19.000 | 41 | 31 | 41 |
| Net yet due tax debt | 2.200 | 4.900 | 4.800 | 10 | 16 | 10 |
| Own capital[1] | 6.200 | 10.000 | 12.400 | 28 | 32 | 27 |
| Total | 22.00 | 31.400 | 46.600 | 100 | 100 | 100 |

Financial structure, producer cooperatives

|  | | | | | | |
|---|---|---|---|---|---|---|
| Short term debts | 828 | 1.073 | 2.049 | 29 | 27 | 29 |
| Long term debts | 1.774 | 1.803 | 4.044 | 62 | 45 | 57 |
| Not yet due tax debt | 35 | 444 | 369 | 1 | 11 | 5 |
| Own capital[1] | 204 | 714 | 663 | 7 | 18 | 9 |
| Total | 2.841 | 4.034 | 7.125 | 100 | 100 | 100 |

[1] Percentage own capital = solidity

Source: Regeringens proposition 1978/79:207

*Table 3. Financial analysis*

Financial analysis, all 18 companies

| | Annual average Million kronor | | | In percentage terms | | |
|---|---|---|---|---|---|---|
| | 1973–1974 | 1975–1977 | 1973–1977 | 1973–1974 | 1975–1977 | 1973–1977 |
| **Funds provided:** | | | | | | |
| Retained earnings | 4.200 | 1.800 | 2.700 | 93 | 34 | 55 |
| New issues of shares for cash | 100 | 300 | 200 | 2 | 6 | 4 |
| Increase of long term debts | 200 | 3.200 | 2.000 | 4 | 60 | 41 |
| Total funds provided | 4.500 | 5.300 | 4.900 | 100 | 100 | 100 |
| **Funds used:** | | | | | | |
| Investment in fixed capital etc | 2.300 | 3.900 | 3.200 | 51 | 74 | 65 |
| Investment in financial assets | 300 | 700 | 500 | 7 | 13 | 10 |
| Increase of working capital | 1.900 | 700 | 1.200 | 42 | 13 | 24 |
| Total funds used | 4.500 | 5.300 | 4.900 | 100 | 100 | 100 |

Financial analysis, producer cooperatives

| | 1973–1974 | 1975–1977 | 1973–1977 | 1973–1974 | 1975–1977 | 1973–1977 |
|---|---|---|---|---|---|---|
| **Funds provided:** | | | | | | |
| Retained earnings | 545 | 95 | 275 | 92 | 14 | 43 |
| New issue | 33 | 4 | 16 | 6 | 1 | 2 |
| Increase of long term debts | 12 | 581 | 354 | 2 | 85 | 55 |
| Total funds provided | 590 | 680 | 645 | 100 | 100 | 100 |
| **Funds used:** | | | | | | |
| Investment in fixed capital etc | 277 | 440 | 375 | 47 | 65 | 58 |
| Investment in financial assets | 87 | 180 | 143 | 15 | 26 | 22 |
| Increase of working capital | 226 | 60 | 127 | 38 | 9 | 20 |
| Total funds used | 590 | 680 | 645 | 100 | 100 | 100 |

Source: Regeringens proposition 1978/79:207.

# References

Abrahamsson, B, *Exemplet Jugoslavien*, Stockholm, AWE/GEBERS, 1976.

Alchian, A A, "The Basis of Some Recent Advances in the Theory of Management of the Firm", *Journal of Industrial Economics*, pp 30–41, November 1965, repr in W Breit and H M Hochman, eds, *Readings in Microeconomics*, 2nd ed, NY, Holt Rinehart and Winston Inc, 1971.

Alchian, A A, "Some Economics of Property Rights", *Il Politico*, 30:816–829, 1965, repr in A A Alchian, *Economic Forces at Work*, Indianapolis, Liberty Press, 1977.

Alchian, A A, "Corporate management and Property Rights", in H Manne ed *Economic Policy and the Re-gulation of Corporate Securities*, Washington: American Enterprise Institute, 1969.

Alchian, A A and W Allen, *University Economics*, 3rd ed, London, Prentice/Hall International, Ind, 1974.

Alchian, A A and H Demsetz, "Production Information Costs, and Economic Organization", *American Economic Review*, 62:777–795, December 1972.

Alchian, A A and H Demsetz, "The Property Rights Paradigm", *Journal of Economic History*, 33:16–27, March 1973.

Cheung, S, "The Structure of a Contract and the Theory of a Non-Exclusive Resource", *The Journal of Law and Economics*, 13:39–70, April 1970.

Cheung, S, *The Myth of Social Cost*, London: The Institute of Economic Affairs, 1978.

Furubotn, E G, "Toward a Dynamic Model of the Yugoslav Firm", *The Canadian Journal of Economics*, 4:182–197, May 1971.

Furubotn, E G and S Pejovich, "Property Rights and Economic Theory: A Survey of Recent Literature", *Journal of Economic Literature*, 10:1137–1162, September 1972.

Furubotn, E G and S Pejovich, "Property Rights, Economic Decentralization, and the Evolution of the Yugoslav Firm, 1965–1972", *The Journal of Law and Economics*, 16:275–302, 1973.

Furubotn, E G and S Pejovich, "Property Rights and the Behavior of the Firm in a Socialist State: The Example of Yugoslavia", *Zeitschrift für Nationalökonomie*, 30:431–454, Winter 1970.

Knight, F H, *Risk, Uncertainty and Profit*, Chicago and London: University of Chicago Press 1971.

Manne, H G, "Mergers and the Market for Corporate Control", *Journal of Political Economy*, 73:110–120, April 1965.

Meade, J, "The Theory of Labour-Managed Firms and of Profit Sharing", *Economic Journal*, 82:402–428, March 1972.

Moberg, R, *Ekonomiska Föreningar*, Uddevalla, Liber 1966.

Pejovich, S, "Towards a General Theory of Property Rights", *Zeitschrift für Nationalökonomie*, 31:141–155, Spring 1971.

Regeringens proposition 1978/79:207, "Om förvärv av aktier i Södra Skogsägarna AB och Norrlands Skogsägares Cellulosa AB".

SIND 1976:1, *Industrin i Sverige, Utveckling till 1980*.

SOU 1979:11, *Löntagarna och kapitaltillväxten 4. Lantbrukskooperationen – ideologi och verklighet*.

*Statistisk Årsbok* 1978.

Ståhl, I, "Ägarformer och företagsbeteende i den svenska blandekonomin", i *Ägarmakt på avskrivning?*, Stockholm, Studieförbundet Näringsliv och Samhälle, 1979.

Ward, B, "The Firm in Illyria: Market Syndicalism", *American Economic Review*, pp 566–589, September 1958.

Williamson, O E, *Corporate Control and Business Behaviour*, Englewood Cliffs, New Jersey, Prentice Hall, Inc, 1970.

# The Social Costs of Union Gains

BY LEVIS A KOCHIN*

*Unions have obtained higher wages for their members. The cost to society of these higher wages are caused by (1) A misallocation of labor – too little employment at unionized work and too much elsewhere (2) Rent Dissipation – the cost of competition for union jobs and of efforts to create and destroy unions (3) Rigidity – Union rules cause an increase in the rigidity of work practice and of wages. A low bound estimate of the social cost of unionism allowing only for the first two categories finds unions had a social cost of $ 58.5 Billion in the U.S. in 1979.*

## I. Introduction

Since the publication of the classic paper "Monopoly and Resource Allocation", [Harberger (1955)], many estimates of the social loss deriving from monopoly in product markets have been made. Fewer attempts have been made to estimate the costs imposed on the U S economy by the monopoly practices of trade unions [Rees (1963), Johnson and Mieszkowski (1970) and Diewert (1974a, 1974b)]. This paper corrects some of this imbalance. Even if a narrow definition of unions is made, a low bound estimate of the social costs of unionism of 2.75 % of the Gross National Products is measured for the United States.

The sources of the social costs of unionism can be usefully placed in three categories: (1) the distortions imposed by the lowered employment of union labor, (2) the expenses incurred in efforts to form, maintain and obtain entry into unions as well as the costs incurred in efforts to obstruct, destroy, harrass or otherwise hinder the efforts of trade unions to maintain wages above those that would exist in the absence of union monopoly, (3) the barriers to progress and to the efficient use of resources which are inevitably entailed by union work rules.[1]

The first two of-these-costs can be located on the diagram, which is similar to that used by Harberger in 1955. The triangle CEF in the diagram represents the net social loss to the economy because workers are employed elsewhere whose marginal social product would be higher in the unionized employment than in other employments. It represents the difference between: (1) a factor demand curve which reflects the marginal

value product of labor in unionized employment, and (2) a supply curve of labor to the unionized industry, which, in the absence of the union, would reflect the alternative product of that labor outside the unionized industry. An estimate of these costs is presented in Section II.

The rectangle ACFG represents the gross gain from membership in their union to those who remain employed in the unionized industry. I will estimate these gains in Section III of this paper. To the extent that property rights in union jobs are ill-defined, the competition of workers to become union members will use resources at least as large as these gross gains. If property rights in union jobs are well defined, these resources will be used in struggles to organize and maintain unions.

The losses imposed by reduced flexibility of management in the unionized sector are more difficult to reckon. These losses, together with those which come about because of the reduced flexibility of nominal wages in unionized industries and the consequent greater fluctuation in real output as aggregate demand changes unexpectedly, are discussed in Section IV.

## II. Derangement of the Stock – The Magnitude Involved

As economists have known for the last two hundred years – and as they have come to measure over the last twenty years – monopolies of any form will lower gross output for:

"Every derangement of the natural distribution of stock is necessarily hurtful to the society in which it takes place; whether it be by repelling from a particular trade the stock which would otherwise go to it, or by attracting

5

Wage

Supply of
Labor to
Union
Sector

Union Wage    A                                    C

          Cost of

Wage if no   B  - - - - - - - - - - - - - -  D  Cost of      E
Union                                          Shifted
          Search For                           Employ-
                                               ment
          Union Jobs

Non-Union    G
Wage                                           F

                                               Demand for
                                               Labor in Union
                                               Sector

                                               EMPLOYMENT
          Employment    Employment
          Given the     With no
          Union         Union
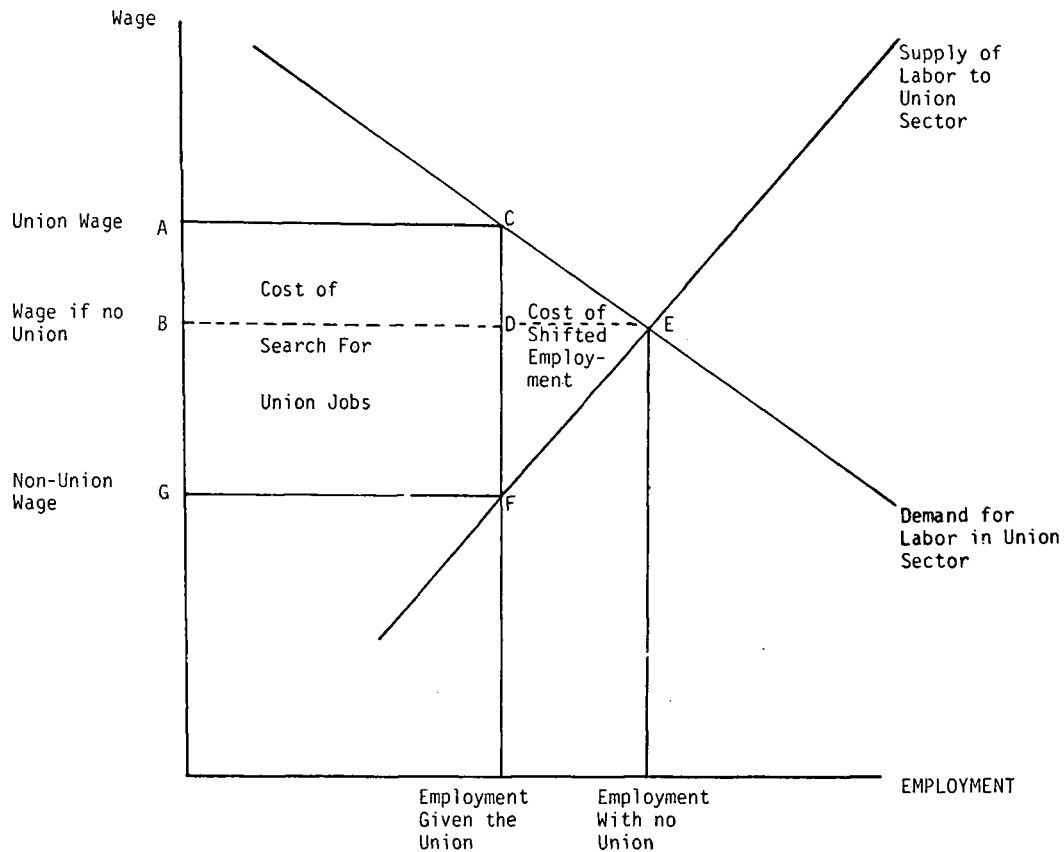
DIAGRAM

THE SOCIAL COST OF UNIONISM

toward a particular trade that which would otherwise not come to it." [Smith (1937, p 597)]

To the extent that unions raise wages without raising the productivity of a given work force, employment decreases in the unionized sector and increases elsewhere. This will cause the stock of labor to be deranged from its most profitable employment.

The questions are how much are wages raised by unions and how large is the resulting alteration in employment patterns. These are not new questions. A standard source in addressing these issues remains Lewis (1963) which summarized all the previous estimates (many of which were the products of the doctoral dissertations of his students)

and provided much new of his own. The weight of Lewis' scholarship was one of the forces which have persuaded others to avoid this area. The attraction of newer labor economic questions has been more important in part because:

". . . unionism in its present form has been quite secure since the early 1950's; its existence is not a serious matter of public policy as it was prior to that time. Hence, there is little call (i e, no funding) for a judgment on the question of whether unionism is a good or bad thing.' [Johnson (1975, p 23)]

The distortion caused by decreased employment of labor in unionized industry is measured by the triangle CEF in the Diagram. This welfare effect will tend to be small because misuse of resources has much smaller effects on output than not using

them at all. The welfare consequences of the al-
teration in the employment of labor which occur
as a consequence of union activity are larger the
more uneven the union effect. The welfare effects
of wage changes increase exactly as the square
of the proportional wage effect if the employment
effect of unions rises linearly with the wage effect.

Lewis's results which are close to those of the
latest studies show:[2] [Lewis (1963, p 9)]:

1) Unions with relative wage rises of 25 %
   which comprise about one-fifth of all union
   members or 5 % of the U S labor force.
   These unions are largely "referral unions"
   in that the union refers new employees to
   the employer.
2) Unions which have achieved wage rises of
   10 % which constitute one-half of all union
   members or about 12 1/2 % of the labor
   force.
3) Unions which have left wages where they
   would have been in the absence of unio-
   nism which comprise 30 % of all union
   members or about 7 1/2 % of the labor
   force.
4) Non-union employment where wages have
   fallen 4 %.

In order to estimate the effect of these wage
rises on employment in unionized industries, it
is necessary to supply an estimate of the elasticity
of demand for union labor. I will use a long-run
demand elasticity for labor of $-1$ which is an ab-
solute minimum for the long-run demand for
labor for a wealth maximizing union. A complete
estimation of the losses also demands an estimate
of the elasticity of supply of labor to the union
sector. A supply elasticity estimate which is con-
sistent with Lewis' estimates of the wage effects
of unionism on non-union wages would be $+4$.

To the extent that a union is able to solve its
internal problems of decision-making it will be-
have so as to maximize the wealth of its owners
who in most cases are largely its members. A we-
alth maximizing union would in a static world
set wages where the marginal revenue from ad-
ditional sales of labor equals the marginal cost of
that labor. In general, the marginal cost of the labor
will be equal to or above the non-union wage.
It will be higher than the non-union wage to the
extent that the union employs monopsony power
against potential members. It will be equal to the
non-union wage if it does not employ such mo-
nopsony power. A simple monopoly union will
set a wage such that:

$$\text{Wage}_{\text{Union}} = \frac{\text{Wage}_{\text{Non-Union}}}{(1+\frac{1}{\epsilon})} \text{ Long-run}$$

To the extent that unions are maximizing the
continous flow of future monopoly profits the es-
timates of union effect on wage costs imply long-
run elasticities of demand for union labor of $-5$
for unions which raise wages by 25 % and $-11$
for unions which raise wages by 10 %. It should
be remembered that these estimates are estimates
at current wages of the long-run elasticity of de-
mand for *union* labor. They are *not* estimates of
the elasticity of demand for labor. One of the most
important possibilities of substitution in the long-
run comes through employment of non-union la-
bor. The estimate of the elasticity of demand for
labor at $-1$ is a low estimate of the true elasticity
and the estimated cost of the lowered employment
is also a low estimate.

This wealth maximizing model should be dis-
tinguished sharply from a temporary profit or
union rent maximizing model. The wealth max-
imizing model assumes that the unions' decision
makers take account of the long-run effects of
short-run gains.[3] To this extent it circumvents
the curious limit pricing constructions of Bain
(1949) and Modigliani (1958). This sort of model
is becoming common in discussions of product
monopoly [for example, Stiglitz (1976)].

The model of the maximizing union has been
explicitly rejected by the leading authorities in the
economic study of unions. The basis for this re-
jection has been institutional and not systematic
examination of evidence. To the extent that they
reject the wealth maximizing model as well, they
are explicitly or implicitly assuming that the cost
of coordinating unions is high enough that im-
portant wealth opportunities which could be ob-
tained by wealth maximizing behavior are igno-
red. The models of utility maximization with both
employment and wages as goods that have been
substituted by Atherton (1973) and Rees (1977)
are generally freer of refutable implications than
is a wealth maximizing model.[4]

Dunlop (1944) argued for wage bill maximiza-
tion – that is, profit maximization without refe-
rence to the fact that union labor has an alternative
cost. Simons (1948) predicted that unions would
set wages just high enough to let all existing mem-
bers find work but will completely shut out others.
The Dunlop model implies that some unions will
lower wages – which has not been observed. Si-

mons' model predicts that no new members will be recruited – but most unions recruit new members.[5]

One implication that can be drawn from a wealth maximizing model of the union is that any interference with the future benefits present workers can derive by letting additional workers into the unions will cause wages to rise and the unions share of its market to start shrinking. If this implication is true, the wealth maximizing model of union behavior has passed a test and can be used with more confidence.

One such event was the removal in the 1960s and early 1970s from referral unions – principally in the building trades – of much of their power to select new members from among the sons and other relatives of the existing members. This power was seized by the courts because property rights in jobs was being inherited in a racially biased manner. If intergenerational transfers are operative, union members would tend to treat the future earnings of new members (that is, their children) as if they would earn them themselves. For if parents are giving funds to children, a dollar to the child must on the margin be as valuable to the parent as a dollar to the parent.[6]

The upshot of this change in property rights was a surge in the wages of workers in the building trades. The building trades raised wages above the wealth maximizing point toward a current profit maximizing wage. From 1968–1971 the average wages of union journeymen building trade workers rose 34.1 % while wages in durable goods manufacturing rose 18.8 %. Soon thereafter the market share of union construction firms started dropping.

Union officials in the building trades consistently opposed the wage surge. In fact, they ultimately obtained the assistance of the United States government in suppressing their members by obtaining government regulation of building union wages. To the extent that the property rights of union officials were not under attack, union officials had no incentives to shift to a more present-oriented view of when the union should extract its rents. The bulk of union activity vis-a-vis the rest of society can be explained as an attempt to maximize the wealth of the members.[7] It is unlikely that the proportion of union generated wealth by union officials is lower than the proportion owned by corporate officials. The discount on "Closed End Funds" and the premia on takeovers are evidence that a non-trivial proportion of corporate wealth is owned by management.[8]

Most of the losses from too low employment in the union sector are in craft unions where the wage premia are largest. The Teamsters have also succeeded in raising wages of long haul truck drivers by 30–40 % [Moore (1978)]. The only industrial union which has been reported to have raised wages by any large fraction is the United Mine Workers. Most of these referral unions – as Ashenfelter refers to them – have probably succeeded in raising wages by 25 % over non-union wages or by about 20 % over the level they would have reached if there had been no union. I assume that the average craft unions have raised wages by 20 % and that the elasticity of demand in these unions is -1 – then they will have lowered employment by 20 % as well. The resultant loss in welfare would be about 2 % of the wage bill in these industries.

Assuming a linear demand curve for labor, the cost of reduced employment in referral unions, which corresponds to the triangle labeled CED in the Diagram, is:

Cost of Reduced Employment
in Referral Unions is:

$$\simeq \frac{1}{2} \times 20\% \times 20\% \times \text{Wage Bill}$$

$$= 2\% \text{ Wage Bill in referral unions.}$$

The wage bill in referral unions is a little more than their total share in employment–wages in craft unions are higher than the average in the labor force. Since total wages are about 20 % of GNP, the total wage bill in craft unions is about 3.5 % of GNP so the total costs of reduced employment in referral unions are about .07 % of GNP.

The losses in the industrial sector are smaller since the rise in industrial union wages is about 6 %. Using the same procedure the losses due to too small employment in industrial unions would be .18 % of the total wage bill in these industries or .015 % of GNP.

The loss due to excess employment in the non-union sector which corresponds to the bottom part of the triangle CEF is about .04 % of GNP. Since wages in the non-union sector are reduced about 4 %, loss due to excess employment is about .08 % of the non-union wage bill but this comes to .04 % of GNP as non-union wages are about one-half of total GNP.

So the total loss due to reduced employment in the unionized sector and increased employment

in the non-unionized sector in the U S is about .165 % of GNP of about $ 3.5 billion in 1979. The large losses are not the consequence of displacement of resources. Misusing resources is less costly than wasting them.

### III. Costs of Competition for Union Wages

Union wages are on the average 10 % higher than non-union wages to others of similar skills. If no well defined property rights existed in job openings in unions, and if these jobs were distributed at random to the first suitable unemployed applicant, then it would pay young workers to spend, on the average, one year or more searching for new union jobs. The process of search must in any case have present value costs as large as the present value of the benefits.

The rate of return to investment in other forms of human capital is less than 10 %. Foregone wages from one year's search as a young person will be less than the average of that person's average lifetime yearly earnings. A year's search as a young person that ended in a new job with 10 % higher earnings would therefore yield a higher rate of return than other forms of human capital investment. This excess search would then appear as a rise in unemployment. This excess search would, on the margin, dissipate the entire excess wages and new union members would derive no benefits from the existence of the union. [See Harberger (1971) and Eaton and Neher (1975).]

Each union member would have a "de facto" property right in a job, which once he had attained it, would be valuable to him. On the other hand, once the initial generation of union members had died or retired, one might observe no individuals whose discounted value of lifetime earning was higher because of the existence of unions. In fact, the lifetime earnings of each individual would be lower because lifetime earnings in the non-union sector would be decreased by the competition of the additional labor displaced by the union sector and the discounted value of unionized workers' wages would be equal to the discounted value of non-unionized workers' wages.

This dissipation might appear, however, in the form of a rise in the average quality of hourly employed workers. This is particularly likely if, as in many industrial unions, the employer has complete control of hiring. There is some evidence that such a rise in average quality of membership has taken place. In a number of measurable ways – most particularly education – union members

tend to exceed similarly occupied non-union employees. The excess of union wages over those available elsewhere tends to reduce voluntary quits and tends to increase the average experience level of workers. Either force would lead to an increase in the quality of labor and would lead to an increase in the demand function for labor.

Any union has a strong incentive to eliminate the dissipations by specifying who is to enter the union. In large part this is the reason why unions with large wage effects specify who is to enter the union. It is the reason why unions with strong wage effect become "referral" unions. The union either arranges for that property to be sold to new members or provides procedures whereby present members can will their spots to their heirs. Cash sales of union memberships at prices which anywhere near reflect the present value of the excess wages are unusual (such prices in the United States today would be in the tens of thousands of dollars for a referral union membership). Sales of memberships on credit would be reflected in a steeper pattern of wage rise in union than in non-union jobs. The opposite seems to be observed – the rise in wage with experience seems to be less on union jobs.

Some possible explanations of this apparent contradiction are: (1) seniority: new workers pay for their jobs by taking less frequent work. Comparisons of the earnings of union members would find a much steeper rise with experience. (2) A selection effect is operating. The older and hardest working employees are promoted to supervisory positions or into the union leadership and leave the union membership. (3) The higher compensation to the more senior members takes the form of fringe benefits in union employment more frequently than elsewhere.

The most frequent way of allocating new jobs in referral unions is by inheritance. This procedure has the defect of usufruct property in that the only way for one's children to collect this inheritance is by continuing in the parental occupation. To the extent that fathers and children have different comparative advantages, this procedure will be costly. But this would be limited since parents would raise their children to follow in their footsteps.

Another form of competition that reduces union wage gains is the "speed up" or other employer techniques for lowering costs or increasing output at the expense of the on-the-job amenities of employees. Unionized employers can "get away"

with this in the market even in the long run to the extent that union wages exceed non-union wages. The non-unionized employer faces a rather flat supply curve and a reduction in on-the-job amenities would in the long run cost him in terms of extra pay. The non-union employer will choose the package of on-the-job amenities that his employees are just willing to pay for on the margin.[9]

But even the most perfect mechanism for specifying ownership of the union will not eliminate competition for union wages as a source of social cost. Union positions can be obtained by forming unions.

This temptation is not unknown to unionized workers or to their bargainers. As a result union contracts are complex documents which often specify in great detail the working conditions of the employees. There is evidence indicating that, nonetheless, the working conditions of unionized workers are worse than those of non-unionized workers. Unionized workers report themselves considerably less satisfied with the non-wage conditions of their employment than do non-unionized workers. [See Borjas (1979).] Even if the negotiations succeeded in specifying conditions in enough detail to prevent on balance any reduction in amenities, the cost of the greater inflexibility imparted to the organization may be significant. (See Section IV below.)

To some extent this area, ACFG in the Diagram, may be an underestimate. To the extent that identifiable groups lose, they will have an incentive to spend resources to resist union organizations. In the short run employers of union labor are the main losers. In the longer run the bulk of the losses will be borne by those workers whose wages fall, when unions gain. The total gains to union members and officers from the existence of unions is likely to be smaller than the return that could have been obtained if the same effort had been used for productive purposes.

Considering all those industries subject to being unionized the area of potential union gains can be regarded as property whose ownership is not clearly set out. If there are many potential claimants, and costs of negotiating among them are high, the entire sum subject to dispute will be exhausted in the expenditures of the potential unionist, their employers and their employers' customers.

The initial formation of a union is not normally costless. Organizers, workers and others must devote human and material capital to this purpose. Also, employers will direct some of their efforts to prevent the formation of unions. The net result is that the costs incurred in forming and preventing the formation of unions may be much larger than the capitalized value of the total gains to successful unionists. In the limit, the cost will be equal to the entire amount which would be gained by the formation of unions, including the amount which could be gained by unionists of industries which are not in fact unionized, since if more is spent at least some participants can certainly gain on average by withdrawing from the struggle.

This suggests that not only are continuing losses due to unionism caused by excess leisure and excess employment in other industries, but other losses as well which must be included if a full accounting is to be made of the costs of unionism. These additional losses may occur as unions are formed or may continue as the struggle continues to obtain wages in excess of the competitive rate and to lower union wages. One result of this struggle will be less output than could have been obtained with the inputs used if somehow monopoly unionism could be made to disappear without a struggle.

If the costs of fighting unions are ignored, the amount of resources used in competing for union jobs will be equal in value to the total extra wages collected by unionists. So including both continuing costs and organizing costs:

```
Cost of Competition
  for Union Jobs is:
```

$$= \frac{W_{Union} - W_{Non-Union}}{W_{Union}}$$

```
x Union Wage Bill
```

Using the Lewis estimate that union wages are made 10 % higher and non-union wages are made 5 % lower, that 25 % of all wages are earned by unionized workers and the wages are about 70 % of GNP

```
Cost of Competition for Union Jobs is:
```

$$\approx .15 \times .25 \times .7 \times GNP$$

$$\approx 2.6\% \text{ of GNP}$$

In 1979, this would be about $ 55 billion. Then the losses from the expense of obtaining union employment are fifteen times the total losses from

the displacement of union labor to non-union employment.

How do these costs appear? One element is the costs of operating unions and is reflected in union dues.

In the United States in 1976 the total revenues of labor unions were about $ 5 billion. Another element appears in the costs of strikes and lock-outs. In the U S the total recorded for such strife has never in the last thirty years exceeded 2 % of unionized workers' hours in any given year. More typically it would be 1 % of the available hours of unionized workers. The loss of output is probably about as large as the lost wages. In the U S this would be about $ 4 billion in a usual year. But work stoppages are not the only cost of union organization nor are dues the only cost of maintenance of the union structure. In fact, our estimate is that together these amount to about 20 % of the union gain of $ 55 billion.

Most successful unions are dependent on broad volunteer effort at the inception. I know of no attempt to cost such attempts. In my opinion one of the largest costs has been on the purchase both from the general public and from politicians of a broadly tolerant attitude toward this one kind of monopoly and toward the private force used to enforce it. Many who view monopoly with alarm view complete monopolies of the supply of particular types of labor with equanimity. This attitude has been purchased by a century of effort.

## IV. Unions Efficiency and Inflation

Unions, as we have seen, cause a waste of resources by "deranging the stock" and inducing privately productive, if socially unproductive, struggles over monopoly union wage premiums. But these effects are not the entire story. Unions in general have the effect of formalizing relations at the working place. This formality has several effects. It makes: 1) adjustment to changing circumstances more difficult; 2) workers more confident of management adherence to bargain and more willing to invest in specific human capital; 3) wages less responsive to changes in economic conditions.[10]

A union contract makes jobs into property. It must do this if the rise in wages is to benefit those workers who obtained it. A seniority system for determining who is to be continued at work is necessary if the workers who obtained the union are to gain from the union. The change of jobs into property has much the same effect as rent

control laws which give existing tenants security of tenure. In a world of zero transaction costs, such a change would not affect the use of resources which would still be bid to their most productive uses. [See Coase (1960).] Such complex tenure systems have real costs in a world where transactions are not free. They increase the initial frictional costs of change and thus bias the system toward continuing existing patterns of resource use.

It is notable that those American industries which have made the largest technical progress are in general non-union. In the world of data processing, of electronics, many of the most avowedly liberal companies have remained non-union. This freedom is a near necessity for firms when the methods of making and operating their products change radically every five years or so. I believe that these effects are important but they are difficult to quantify, and I cannot make any estimate of their importance.

One detailed examination of the effects of the unionization of British coal mining during the twenty years before World War I [Pencavel (1978, p 145)] estimates that output efficiency, that is holding all inputs constant, fell 22 % when the coal mines were unionized.

Brown and Medoff (1977) have produced an almost exactly contrary estimate of the effect of unions on productivity based on a regression estimate of output in various American industries. They estimate that unionized workers productivity is about 22 % higher than that of non-union workers. When some unexplained industry dummies are omitted this effect vanishes. The Brown-Medoff estimate implies that unions raise wages no more than they raise productivity. If this were true employers would have no reason to resist unions but could be induced to welcome union organization. The Rosen (1969) and Lee (1978) findings that union effects are larger in fully unionized industries than in others fit the union-monopoly explanation of union wage premiums but not the productivity explanation.

Another aspect of the formalization of the work relationship by unionism is seen in the sharp contraction of wage spreads across workers in a given occupation once that occupation is organized. Their contraction is a consequence of the facts that with the higher union wage there is little way to use low wages as penalties to the less productive workers. This contraction of wage differentials must reduce incentives.

In some countries in Europe during the 1950s and 1960s the union wage lost its previous character as a standard wage. The great bulk of workers were paid wages above those contracted for. This situation is a sign that the unions had lost their monopoly powers, though in many cases they retained extensive influence both as political and economic agents.

Unions have some favorable effects on productivity. The very formality and rigidity which they promote at the workplace offers to workers a protection against the petty tyranny of their immediate supervisors. This protection should be useful to the employer as well. Foremen whose primary rewards come from maximizing counted output and minimizing counted inputs in the area they control will have a strong inducement to work their employees harder than higher management would want and to otherwise exploit them. This is a consequence of the fact that the lowered reputation of the factory as a "good place to work" will be borne not only by that foreman but by the plant or firm as a whole. These effects are likely to last beyond the normal tenure of a manager and if they are not directly measured the net contribution of an oppressive manager to the firm will be less than that which he is credited with.

Unions offer a path by which complaints can be processed. In this sense a union can be thought of as an external personnel department. One prediction is that unionized firms will have smaller personnel departments than non-unionized firms. These services, however, can be provided by purely voluntary unions. There are fewer services which it is easier to exclude free riders from than the processing of grievances (Reynolds, 1977). But monopoly unionism benefits all those who are allowed to remain in the workplace. The externality argument for compulsory unionism is an argument for compulsory collection of funds for a product-monopolization which is a private good and a public bad.

## V. Summary

Adding together the more easily measured losses from monopoly unionism, we find that these are about 2.75 % of GNP. This estimate is dependent on the assumption that the directly productive net effects of unionism are small or negative. It is an underestimate of the costs as we have minimized the losses from unionism by using a low estimate of the amount of labor displaced by hig-

her union wages. If we had used the higher elasticities of demand for union labor consistent with the wealth maximizing hypothesis of union behavior, the estimate of the loss from displacement of labor would be five or ten times larger. We also excluded from consideration the losses that occur as monopoly unionism is resisted. Nonetheless, these estimates add up to $ 58.5 billion/ year in 1979 which is about half the size of the U.S. Defense budget.

## Footnotes

* University of Washington, Seattle. Yoram Barzel, Dan C Heldman, Masanori Hashimoto, John Hause, John Pencavel, Charles Stuart and Spencer Wedlund have contributed to this piece. Seminars at Yale, Princeton, Stanford and the University of Washington have purged errors and added items of interest. The National Right to Work Foundation funded this study. The work was completed while I was at the Hoover Institution.

1. The approach in this paper is an extension of Tullock (1967). For a discussion of the historical antecedents of Tullock's arguments see Kochin (1980). Krueger (1974) and Posner (1974) supply estimates of losses due to the competitative pursuit of artificial scarcity rents by product monopolists. Harberger (1971) applies the Tullock approach to labor market distortions in developing countries.

2. Since Lewis (1963) a fair amount of effort has been devoted by economists to estimating the wage impact of unions. Weiss (1966) and others who have used ordinary least squares and individual data have generally found higher estimates of the union wage premium than did Lewis. An advantage of micro data was that it could be used to obtain interesting findings, such as the union differentials are larger for blacks than whites except in construction [Ashenfelter (1973)]; lower for better educated workers [Johnson and Youmans (1971)]; lower for the inexperienced [Neuman (1977)]; which could not have been done in any other way.

Ashenfelter and Johnson (1972) argued that unionism would attract workers with long prospective tenures at their jobs since the individual worker would benefit more from a union the longer he expected to work at a job. Such employees would be paid well anyway as a wage premium is a entirely sensible technique for keeping down costly turnover. When they took account of this interdependency the union wage premium diminished.

Later studies which use individual data and take account of this interdependency find substantial wage premia. See Schmidt and Strauss (1976) – 10.5 % and Neumann (1977) – 9 %. These are close to the Lewis estimates as Neumann remarks (p 17).

3. Employment is set a quantity at which the unions long run marginal revenue equals the non-union wage if the demand for the services of the monopoly is expanding at the rate of interest. If demand expands at a rate below the discount rate it will pay to "milk" the firm so as to allow the monopoly to vanish. The union wage would lie above the long-run profit maximizing wage.

4. Reynolds (1978) draws a false prediction from Atherton's model that rises in alternative wages should have no effect on union wages and employment. Rent and wealth maximizing models have the prediction that a rise in wage elsewhere should lead to a rise in union wages and a fall in union employment if the demand for union labor has remained constant. The estimates of Mitchell (1978) and others which show a strong effect of other wages on union wages can be interpreted as supporting rent and wealth maximizing models and falsifying Atherton's model.

5. As Rees (1977, pp 49-51) points out, unions have seldom committed suicide in the fashion predicted by Simons (1948). The building trade maintained their market shares in urban construction intact for 50 years until the early 1970s. Unions which have obtained for present members the net gain from admitting new members by nepotism, pensions, seniority, or (rarest of all) direct sale of job right have no incentive to commit suicide.

6. For a discussion of the ways that gifts tie generations see Becker (1974). Some of the unions which were de-segregated attempted to ward off the courts by pointing out that their gifts of membership to their children were no different from other inheritance. These pleas were not accepted even in a case in which some of the sons were black. [Gould (1977, p 289).]
Job seniority is judicially regarded as non-transferable property (ibid, p 288).

7. Some other implications of wealth maximizing behavior to Unions can be seen worked out and tested in Powell (1973), an unpublished dissertation on the A.M.A.

8. These discounts are 20 %-30 % at the moment.

9. The estimates of the wage premium to union members are net of the effect of unionism on the composition of the labor force. The increase in the productive attributes of unionized workers involves a loss to the extent that the employment of skills is deranged.

10. Some of these costs are imposed on unionized workers; unionized workers are far more likely to be laid off than are other workers and far less likely to have their wage lowered in a recession. Hashimoto (1975).

# References

Ashenfelter, O, (1972), "Racial Discrimination and Trade Unionism." Journal of Political Economy, Vol 80, pp 435-65.

Ashenfelter, O and Johnson, G, (1972), "Unionism, Relative Wages and Labor Quality in U S Manufacturing Industries." International Economic Review, Vol XII, pp 488-508.

Atherton, W N, (1973), Theory of Union Bargaining Goals. Princeton: Princeton University Press.

Becker, G, (1974), "A Theory of Social Interactions." Journal of Political Economy, Vol 82, pp 1069-93.

Boskin, M J, (1972), "Unions and Relative Wages." American Economic Review, Vol LXII, pp 466-72.

Borjas, G, (1979), "Job Satisfaction, Wages and Unions." The Journal of Human Resources, Vol XIV, pp 21-40.

Brown, C and Medoff J, (1978), "Trade Unions in the Production Process." Journal of Political Economy, Vol 86, pp 355-78.

Coase, R, (1960), "The Problem of Social Cost." The Journal of Law and Economics, Vol III, pp 1-45.

Diewert, W E, (August 1974), "Unions in a General Equilibrium Model." The Canadian Journal of Economics, Vol VII, pp 475-95.

, (1974), "The Effects of Unionization on Wages and Employment: A General Equilibrium Analysis." Economic Inquiry, Vol XIII, pp 319-39.

Dunlop, J, (1944), Wage Determination Under Trade Unionism. New York: Macmillan.

Eaton, B and Neher, P A, (1975). "Unemployment, Underemployment and Optimal Job Search." Journal of Political Economy, Vol 82, pp 355-75.

Flanagan, R J, (1976), "Wage Interdepence in Unionized Labor Markets." Brookings Paper on Economic Activity, 1976:3, pp 635-82.

Freeman, R B, and Medoff, J L, (1979), "New Estimates of Private Sector Unionism in the United States." Industrial and Labor Relations Review, Vol 32, pp 143-75.

, "What Do Unions Do? A Progress Report." Mimeo. Harvard University Department of Economics. (1977)

Gould, W B, (1977), Black Workers in White Unions, Ithaca: Cornell University Press.

Hall, R E, (1975), "The Rigidity of Wages and the Persistance of Unemployment." Brookings Papers on Economic Activity, 1975:2, pp 301-50.

Harberger, A C, (1955), "Monopoly and Resource Allocation." American Economic Review, Vol LV, pp 77-87.

, (1971), "On Measuring the Social Opportunity Cost of Labor." International Labor Review, No 103, pp 559-79.

Hashimoto, M, (1975), "Wage Reduction, Unemployment and Specific Human Capital." Economic Inquiry, Vol XIII, pp 485-504.

Johnson, G E, (1975), "Economic Analysis of Trade Unionism." American Economic Review, Vol LXV, pp 23-28.

334  Levis A Kochin

Johnson, H and Mieszkowski, P, (1970), "The Effect of Unionization on the Distribution of Income: A General Equilibrium Approach." *Quarterly Journal of Economics*, Vol OO, pp 539–61.

Kochin, L A, (1980), "Monopoly Profits and Social Losses." *Research in Law and Economics*, Vol 2, pp 201-12.

Krueger, A O, (1974), "The Political Economy of the Rent Seeking Society." *American Economic Review*, Vol LXIV, pp 291–303.

Lee, L F, (1978), "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables." *International Economic Review*, Vol 19, pp 415–31.

Leontief, W, (1946), "The Pure Theory of the Guaranteed Annual Wage Contract." *Journal of Political Economy*, Vol 54, pp 76–79.

Lewis, H G, (1963), *Unionism and Relative Wages in the United States*. Chicago: University of Chicago Press.

Martin, D L, (1978), *An Ownership Theory of the Trade Union*. Mimeo. University of Miami.

Mitchell, D, (1978), "Union Wage Determination: Policy Implications and Outlook." *Brookings Papers on Economic Activity*, 1978:3, pp 537–82.

Moore, T G, (1978), "The Beneficiaries of Trucking Regulation." *The Journal of Law and Economics*, Vol XXI, pp 327–43.

Neumann, G R, (1977), *Union Wage Differentials and the Decision to Join Unions*. Chicago: University of Chicago Press.

Pencavel, J H, (1977), "The Distributional and Efficiency Effects of Trade Unions in Britain.', *British Journal of Industrial Relations*, Vol XV, pp 137–56.

Perloff, J M, (1976), *The Wage Change Process in the Construction Industry*. Unpublished Ph D Thesis, M I T.

Posner, R A, (1975), "The Social Cost of Monopoly and Regulation." *Journal of Political Economy*, Vol 83, pp 807–27.

Powell, J H, (1973), *A Theory of Trade Union Behavior Applied to the Medical Profession*. Ph D Dissertation, University of Washington.

Rees, A, (1963), "The Effects of Unions on Resource Allocation." *Journal of Law and Economics*, Vol VI, pp 69–78.

, (1977), *The Economics of Trade Unions*. Chicago: University of Chicago Press.

Reynolds, M O, (1977), "Labor Unions, Cartel Theory, and Free Riders." *Mimeo*. Texas A&M Univ.

, (1978), "A Cartel Model of Labor Union Behavior." *Mimeo*. Texas A&M Univ.

Rosen, S, (1964), "Trade Union Power, Threat Effects and the Extent of Organization." *Review of Economic Studies*, Vol 36, pp 185–96.

Simons, H, (1948), "Some Reflections on Syndicalism," in *Economic Policy for a Free Society*. Chicago: University of Chicago Press.

Smith, A, (1937), *The Wealth of Nations*. The Modern Library, New York.

Stiglitz, J E, (1976), "Monopoly and the Rate of Extraction in Exhaustible Resources." *American Economic Review*, Vol LXVI, pp 655–61.

Tullock, G, (1967), "The Welfare Cost of Tariff. Monopoly and Theft." *Western Economic Journal*, Vol 4, pp 224–32.

Warren-Boulton, F R, (1977), "Vertical Control by Labor Unions." *American Economic Review*, Vol LXVII, pp 309–22.

# The Supply and Demand for Protection: A Suggestion for a Positive Theory of Democratic Government

BY MICHAEL D BORDO AND DANIEL LANDAU*

## Introduction

With the growth of government intervention in western market or "mixed" economies, economists have been directing increased attention to a positive analysis of government behavior. By a positive analysis is meant a study of the relation between the preferences of those groups (or individuals) controlling the government and government policies; as opposed to a "normative" study of what government policies ought to be to serve the "general good".

However, if the government to be analyzed is democratic and by democratic government is meant to be government "by, of, and for the people", then we would expect government to serve the general interest. Therefore, if major policies which persist over long periods of time *seem* to serve special minority interests, these policies present a major theoretical dilemma for a positive theory of democratic government. This dilemma can be resolved in one of two ways; first, by redefining democracy to mean something less than a government controlled by the governed in their self-interest; secondly, by showing that these policies in fact are a response to the demand of the majority of the electorate and that the view that they serve minority special interests at the expense of the large majority is incorrect.

The first approach is illustrated by a considerable number of books and articles from the field of public choice. Good examples are Stigler's *The Citizen and the State*(1975); *Capitalism and Freedom: Problems and Prospects*, Selden (ed) (1975), and Borcherding *Budgets and Bureaucrats* (1978).

In this paper we propose to take the second route. We provide a re-interpretation of government interventions in markets and other policies based on a perception of government in modern democratic nation states as being principally an expanded "protection agency". Beginning with a discussion of national defense, we argue that much of what modern government does is to protect the lives, property, and permanent income of its citizens from unacceptable variations of both external and internal causes.

We believe our approach can explain the behavior of democratic government – i. e. of political markets – in a manner similar to the analysis of economic markets. Thus, individuals are assumed to be motivated in their demand for government policies by self-interest, and government elected officials – "firms" – operate in a regime of effective competition which forces them in the long-run to supply voter-demanded policies.

The key elements of our approach are developed in the next section. Section III compares our approach to the public choice view. Section IV continues the analysis and demonstrates how the size and complexity of modern government can be traced to various aspects of the market for protection. Section V provides some qualitative evidence in support of our approach from the modern history of democratic states. The conclusion sums up the paper and suggests various possible extensions.

## II

In this section of the paper we explain our approach in terms of the supply and demand for government services. In the process we discuss five major issues: (i) the demand for government services as a demand for protection; (ii) the factors which allow the government to supply many forms of protection cheaper than the market; (iii) citizen control of government through political markets; (iv) the differences between short and long run supply of government services; and (v) the difference between the national and lower levels of government.

Economics views the individual as a wealth or utility maximizer. Given his resources, the individual will endeavor to attain a consumption path which will maximize utility. The consumption path actually attained will depend both on the normal returns to the factors an individual owns and the variations around these normal returns.

Asset and insurance markets allow the individual to reduce the impact of certain types of variation in income on his consumption path at the least cost. Other variations can be dealt with at a lower cost by having government bear the risks.

While some government risk bearing is of the insurance type – (for example disaster and health insurance) – the major component of government risk reduction is a special type which we term "protection". Therefore, in this theoretical section we will confine our discussion to this special form of risk bearing.[1]

Protection differs from insurance in that where insurance compensates the individual (household) for losses, protection reduces the probability of events which cause losses. Thus, the possible losses due to fire can be reduced either by fire-insurance or by building with fire-proof materials. Building with fireproof materials is protection because it reduces the probability of fire.

Certain forms of protection can be provided by the state at lower cost than the market. The fundamental form of protection provided by government is protection against foreign aggression. The government can provide this form of protection at the least cost because it controls the armed forces.[2] The armed forces are a means of coercing -foreign aggressors. The "government" of a country is *in fact* the government because it controls this ultimate means of coercion.[3]

The situation is similar for internal activities of governments; the role of government is ultimately based on its ability to *enforce* laws and law enforcement rests on the coercive powers of the police. Of course, governments also fulfil other functions. However, the private sector could replace the government except where the ultimate sanction of coercion is necessary.[4]

Modern democratic nation states – MDNS – provide their citizens with various forms of protection. Some of these forms of protection are also provided by other types of states, but this paper will discuss only MDNS.

The MDNS can be viewed as a club with the citizens as members and the government as its executive committee.[5] Club members demand

from their executive committee four types of protection for themselves as well as protection for the club. For themselves as individuals, club members demand: protection against external threats (from outside the territory of the club); protection against internal private threats (criminals); protection against internal public threats (abuse by government); protection against certain "unacceptable" decreases in permanent income (e.g. those caused by farm price variations). The club members demand protection of the club as a whole because it is a key input in the production of other forms of protection (e.g. without a nation there is no national defense).

The MDNS have both economic and political markets. Economic markets exhibit the standard properties. In "political markets', the individual club members – as voters – demand government policies which are supplied by "firms" – public officials – and groups of firms' parties.[6] Payment for government services is made to government officials both in terms of the means of achieving office-votes, campaign funds, endorsements, etc, and the returns to office holding, salaries, "perks", publicity etc.

Voter control of the government is exercised through political markets. These political markets of an MDNS are by definition democratic. By democratic we mean that political markets exhibit voter sovereignty and rationality on the demand two approaches: first, their scientific methodolog- on the supply side (in the long-run).

The competition in political markets both between firms-candidates and consortia-parties is quite intense. Office holders have a wide spectrum of goals – re-election, higher offices, more influ= ence in the party and government – for which they compete both with other office holders and with candidates not yet elected. Popularity with the voters is essential to achieving these goals. Therefore politicians must continually be proving to the voters that they are providing the policies the voters desire.

This combination of competition on the supply side along with rationality on the demand side allows voters to enforce their policy preferences on their government in the long run.

Long run equilibrium is attained when voters obtain from government the policies – i.e. protection – they demand at minimum cost. The movement from short-run equilibrium to long-run equilibrium will involve three types of changes: first, elimination of policies (or lack of policies) which benefit special interests rather than the ma-

jority; secondly, voters cease to demand policies which are clearly not in their self-interest or which are simply impossible to achieve (e.g. price controls or zero unemployment); finally, high cost means of providing protection are replaced with low cost means.[7]

Control of the means of coercion and the imposition of high costs of entry and exit from the national club allow the MDNS governments to behave as monopolies over their territory. In contrast, lower level government – state, municipal – must generally behave competitively. These lower levels of government survive because they have a comparative advantage in providing certain forms of internal protection – e.g. police, fire protection, health, etc. Since these lower levels of government behave differently from national government, we limit our analysis to national government.

In sum, our approach is that the national government of MDNS can be better understood as the executive committee of a political club – the nation. The government is chosen in political markets which are effectively competitive (in the long-run). In the club's political markets the club members obtain (in the long-run) the government policies they demand which are, first and foremost, protection of various forms. The suppliers of protection are the elected government officials. Voter – club member – control of government holds (in the long-run) because all elected officials are in constant and multi-dimension competition (between themselves and with potential office holders) both *inside* and between political parties.

## III

In this section we compare our view of the demand for government services as a demand for protection with the major alternative hypothesis from the public choice literature. This alternative view can be stylized as follows; the demand for government services is a demand for public goods and transfers. The latter component is the key determinant in recent decades of the growth of government.

We discuss and compare three aspects of the two approaches: first, their scientific methodological acceptability; secondly, their implications for the supply of government services; thirdly, their consistency with the available empirical evidence.

On methodological grounds there is little basis to choose between the two approaches. Our approach has the slight advantage of explaining go-

vernment growth in recent decades as well as government functions dating from previous centuries from one principle, the protection of wealth. The alternative view uses one principle (public goods) to explain government's on-going function from previous centuries, and another (transfers) to explain its recent growth.

By the more important methodological criteria of refutability, there is little basis to choose between the two approaches. Both require sub-hypotheses to keep them from being excessively general and therefore unrefutable. In section IV of this paper we demonstrate that protection motivation can explain most government activity. Similarly, almost all government policies create transfers, since to *prevent* transfers – irrespective of the motivation for the policy – would require deliberately distributing all benefits in exactly the same way as the costs. The generality of both protection and transfers is due to their close association with the fundamental general economic process of wealth (or utility) maximization; that is, in an uncertain world income stream "X" will be preferred to income stream "Y" if X either has a higher mean (with equal variances) or a lower variance (with equal means). Production and the receipt of transfers are two basic determinants of the mean income stream, while insurance and protection are two basic means of reducing variance in an income stream.

There are various sub-hypotheses or constraints which could be added on to either approach in order to make it testable and refutable. For example, a government policy might be considered as motivated by transfer (protection) considerations only if the transfer (protection) benefits exceeded a certain minimum percentage of the total cost, (e.g. 5 %, 10 %, or 25 %).

When we consider the implications for the supply of government services, the advantages of the protection approach appear to be quite important. *Given* the existence of such threats as fires, criminal attack, and foreign aggression, protection can be a "positive sum game" benefiting most *club members*.[8] In sharp contrast transfers are at best a "zero sum game" and in fact, given the costs of effecting transfers, they are a "negative sum game" with the costs to losers exceeding benefits to gainers.

Since transfers are a negative sum game, rational self-interested maximizers will control their political markets to guarantee that transfers are kept to a minimum. Thus, proponents of the trans-

fer approach seem to be arguing that the concept of democratic government – i.e. government by, of and for the people – is basically an illusion. Considering the importance usually attached to the distinction between democratic and non-democratic governments, this approach could be counter-productive.

The problem is really much deeper. A fundamental postulate of economics is that individuals know their own self-interest, and that firms will find it profitable to supply the self-interest demands of individuals. The transfer theory claims this basic postulate does not apply to the huge and growing government sector. If the elected representatives of households in government can ignore or distort the voters' preferences and tastes, why can't large corporations do likewise? To our thinking the transfer theory is in fact calling in question the foundations of economics.[9]

The transfer theory would appear to be in conflict with the evidence in several important aspects. One is the rationality of voting and deciding whom to vote for. If the government is providing desired (protection) services it may be rational to invest time in the act of voting and in the process of deciding whom to vote for.[10] However, if the government is not controlled by the voters, such use of time is clearly irrational. Thus, we would predict, on the basis of the transfer theory, that people would not vote and would not invest time becoming informed on the issues. Furthermore, we would predict that groups more likely to be rational in their behaviour and which have a higher cost of time, such as the better educated and higher income segments of the population, would vote less and be less informed about the activities of government. In fact these predictions are contradicted by the available evidence in MDNS.

The transfer theory has been around for a long time, but, as Stigler (1975) points out, there has been very little measurement of the size of such transfer. In contrast, the protection approach is new and therefore has not yet had time to be quantitatively tested. Also the transfer theory seems to have difficulties with the evidence from empirical studies of income redistribution. One prediction of many transfer theorists which has been studied empirically is that majority rule voting would appear to allow a small majority of voters to significantly redistribute income from a large (or high income) minority to themselves. This majority could even simply be a coalition of minorities. On this basis, many economists have predicted significant income redistribution towards the poor, the middle class, or groups of minorities which together reach 51 % or more of the electorate. However, various studies of government tax and expenditure policies at the federal level in the U.S. and Canada have found the amount of redistribution to be small.[11]

We feel the drawbacks of the transfer theory are sufficently serious to raise the question, why has the transfer theory survived? We can not provide a full explanation. However, perhaps a partial explanation is that the proponents of this approach have exaggerated the differences between political and economic markets.

Economic and political markets are so different that it is easy to assume they are simply incomparable.[12] In fact, many of what appear to be *basic* differences, between the two types of markets are the result of imperfect specification of the lags, information and transactions costs, services sold, and the nature of suppliers and demanders in political markets. As we pointed out earlier, the political market is a market för policies (not offices) where competition between candidates produces many elements of effective competition. Furthermore, the intermediary firms in political markets are not only political parties, they also include the huge numbers of pressure groups. Finally, economic markets allow different consumers to purchase different products in the same market. Similarly, the political market produces different tax rates and exemptions for different classes of citizens.

In conclusion, for all its intuitive plausability, the transfer theory when carefully analyzed seems to be inferior to the protection approach both in its implications and its consistency with available empirical evidence.

## IV

In this section we examine the five basic categories of protection services that club members demand from their government and we argue that these protection demands can explain most expenditures of MDNS National Governments. The five categories of protection demands are: protection against external threats, protection against internal private threats, protection against internal public threats, protection against "unacceptable" decreases of permanent income, and protection of the club as a corporate body.

## 1. Protection Against External Threats: National Defense

The "production" of national defense by modern armies exhibits economies of scale, is highly capital intensive, and is subject to very rapid technical change. These factors alone would make national defense very expensive. However, the costs of national defense are not limited to armed forces budgets. They also include the cost of subsidized domestic defense industries, the different between budget and opportunity costs of draftees' time, part of the costs of foreign affairs and foreign aid, the cost of maintaining control by the elected civilian government over the armed forces, and more.

## 2. Protection Against Internal Private Threats: Crime

Government expenditure for the protection of club members against criminal elements includes the obvious components of police, courts, and prisons, but these are only a part of government outlays on crime reduction. High schools, recreation facilites, and social welfare programs are thought to reduce crime, and so a share of these much larger outlays must also be assigned to the category of protection against crime.[13]

## 3. Protection Against Internal Public Threats: Government Abuse

The danger to citizens from the best of governments is an ancient and important theme in political science. In general, the monopoly of the government on the means of coercion inside the territory of the club gives it the power to abuse citizens in various ways including imposing confiscatory taxation. Basically, only a branch "i" of the government can protect the individual club member against abuse by another branch "j". This leads to a whole dimension in government size and complexity starting from a bill of rights (enforceable in the courts) continuing with divisions of powers at the level of national governments – the U S checks and balances – as well as with appeals to higher courts, and finally including at a very subtle level the nature of many clubs' government structure. The U S federal structure was, of course, very consciously conceived as a means of limiting the power of the national government. While the direct costs of such complexities are large, the indirect costs are probably enormous. Court cases may be argued and re-

argued through several levels. Significant legislation can be passed by the congress or parliament and rejected by the president or the courts on constitutional grounds.

## 4. Protection Against Unacceptable Decreases in Permanent Income

There are a wide range of government programs to give the individual protection (or simply insurance) against decreases in his permanent (or current) income stream where the decreases are felt by the majority to be both excessive and beyond the individual's control. These programs include insurance type programs (unemployment insurance, social security, government health insurance, disaster relief, etc) and programs to influence supply and prices in various sectors (farm programs, tariffs to protect declining industries). They also include various laws and government regulations intended to protect the consumer (or employee) directly from "dangerous" products (or working conditions).[14]

Each of these income programs is a special case requiring a detailed empirical study. However, some of the more general issues raised by the three categories of such programs mentioned allow more general theoretical answers.

Insurance type programs raise several general issues. Most important, are they purely insurance, or is there a protection component also? What we observe is that most of these programs seem to have protection elements. Unemployment insurance reduces the danger that the unemployed will become a threat to the stability and peaceful economic activity of the club and its employed members. Survivors' insurance has the same benefit with respect to the orphan children of club members. Insurance of a minimum income to the elderly increases the identification of younger adult club members with the club and its stability and security just as the prospect of a company pension increases the employee's identification with the firm. (See discussion of the demand to protect the club as a corporate body below.)

The examples given indicate that the protection benefits from various compulsory insurance programs may be significant and they could easily exceed any inefficiency costs of government instead of voluntary private insurance.[15]

Our theoretical framework is designed to explain the survival in MDNS of government policies which economics has usually found to be harmful to economic welfare. The classic example

of such unwelcome survivors is the tariff, which effectively limits import competition with domestic firms. (We can limit our analysis to tariffs to protect existing industries since tariff protection of new industries is a short-run problem.) As we have learned in our principles courses that tariff protection for existing industries usually reduces economic welfare, tariffs represent a case where our approach runs most counter to an economist's basic intuition. It is helpful to break the problem down into two sub-issues: first, why should one citizen (not owning factors employed in the protected industry) suffer the cost associated with protecting another (who does own factors employed in the protected industry); secondly, why should protection take the form of a tariff instead of a (seeming more efficient) subsidy?

The essential distinction between our approach and the standard analysis is the positive value placed by the individual club member on the survival of the club as the provider of protection to him. The individual will be ready to pay (with higher product prices) to protect and promote the club. Therefore if individual member $X_i$ would want protection for his capital – specific capital to industry I – he will accept as legitimate the demand of individual club member $X_j$ for protection of his specific capital employed in industry J. (The specific capital could be human or non-human, but the human capital in the form of laborers with specialized skills is usually given greater importance.) Thus the club member $X_i$ has two reasons to agree to protection for $X_j$: (i) he wants such protection to be a privilege of club membership and therefore available to him; (ii) allowing such protection to $X_j$ will strengthen $X_j$'s loyalty to the club which protects $X_i$.

*Given* the decision to protect domestic industry J from competitive imports, basic economic theory tells us a subsidy ought to be relatively less costly than a tariff.

However, this elementary analysis assumes the costs of administering protection (direct and indirect) are equal between the two means. In fact, for subsidies both costs are likely to be higher. The direct costs of administering a subsidy are those of knowing how many units of output J each firm in the industry produced and then making the payments. The direct costs of a tariff are those of monitoring flow through ports of entry and collecting payments. The latter will tend to be less since monitoring incoming goods flows is a necessary part of basic protection whereas monitoring output of individual firms *inside* the club

territory is not an essential activity of government.

The indirect costs of a tariff or subsidy are the resulting distortions in the domestic market. A tariff does not require intervention in the domestic market for J; domestic producers are only protected from imports with no commitment to protect them from domestic market competition. A subsidy implies (and may require) protecting domestic producers from domestic market forces in order to guarantee a quantity and price of J which will make imports non-competetive.[16]

Government intervention in internal markets to protect consumers (employees) from "unsafe" products (or working conditions) is clearly an example of protection. Such protection would be in the protected consumers' (employees) self-interest if the gains in utility (expected income) exceed the costs in higher product prices (lower wages). The scope of such regulation has grown enormously in recent years. There is, of course, no way of knowing if the newer regulation will pass the test of long-run survival. However, major examples such as airline safety and nuclear power regulation would appear to fit current voter preferences.

## 5. Protection of the Club as a Corporate Unit

The self-interested individual will value the club as a key input to his own protection. Thus, he will accept government expenditures and interventions to perpetuate and advance his club in a world of many national clubs.[17] He will also favor expenditures and interventions which increase other club members' identification with the club.

Perpetuation of the club implies maintaining its national culture, language, customs, institutions, etc. This results in taxes, subsidies and constraints on individual maximizing behavior which can be quite costly.

The club will tend to be stronger the more homogenous its population since mutual trust tends to come easier in groups of similar people.[18] A growing club population, however, helps the club by allowing economies of scale in national defense and in tariff-protected economic markets. The combined goals lead to measures designed to increase birth rates, improve health, restrict immigration and reduce emigration. Here again, substantial costs may result from programs designed to efficiently promote and protect the national club.

In conclusion, we would list additional types

of government policies which are of a protective nature; however, it would appear that these five categories of protective policies can account for much of the cost of government.[19]

## V

In the preceding sections of the paper we have outlined an approach to explaining the growth and persistence of government in the MDNS. In this section we provide some qualitative evidence in support of our approach from the modern history of a number of democratic nation states.

First, we discuss several examples of the importance of the linkages between protection from external threats, development of nation states, and economic policies.

a. The British Navigation Acts can be viewed as an important form of protection – protection for the English middle class as well as protection for all commercial users of the North Atlantic in the eighteenth century.[20] The British provided a service to their colonies and to other countries by ridding the seas of pirates and protecting the sea lanes from violence. This service was to be paid for by restricting British colonial commerce – between colonies, the mother country and third parties – to British or colonial ships as well as by other policies such as import duties and export bounties.

The absolute defeat of the Dutch navy by the end of the seventeenth century ensured the British navy the complete monopoly right to providing protection. Moreover, some of the clauses of the Navigation Acts which appeared to be to the detriment of the British North American Colonies can be explained as being part of a package of protection against the threat of the French in Canada.[21]

b. The American Revolution in our view largely reflected a major change in both the costs and benefits of providing protection to the thirteen colonies. Before 1763, the British efficiently protected the thirteen colonies from the French in Canada. With the defeat of the French in 1763, the colonies were well able to protect themselves, and moreover, removal of the French threat allowed the British to act as if they had a monopoly on the protection of North America. This was reflected in the British attempt to increase the colonists' contribution to the British Empire after 1763 with new taxes and greater enforcement of the less popular statutes of the Navigation Acts. To the colonists the reduced benefits of protection

(there being no immediate foreign threat) outweighed the costs and we can view the revolution as a rational switch to a more efficient and *more easily controlled* protection agency.[22, 23]

c. In Canadian history, two elements stand out as the pillars of an effective protective system for the Canadian (British North American) club: securing a national transportation system; and the tariff (the National Policy). Confederation in 1867 can be viewed as an arrangement whereby French Canadian language and cultural rights were to be protected in exchange for support of the British North American club against possible encroachment by the U S club.[24] To ensure effective protection of the club, extensive resources were committed to the construction of a national railway, the Canadian Pacific Railroad, and it was imperative to the "Fathers of Confederation" that this railroad only run through Canadian territory, regardless of the cost.[25] To build the railroad, the embryonic Canadian government extended generous land grants, subsidies and other privileges to the railroad.[26] As well, the national tariff policy, established in 1879, was designed to encourage (inefficient) east-west (intra-Canadian) trade to the detriment of (the more efficient) north-south (U S-Canada) trade.[27] An alternative explanation, consistent with our view was provided by the Dean of Canadian Economic History, Innis, who argued that the national tariff was primarily a revenue tariff established to generate the substantial funds to subsidize the building of an (inefficient) national transportation system. Such a system Innis argued was crucial to the protection and survival of the Canadian club.[28]

Secondly, we argue that a large number of major government interventions in modern economies can also be regarded as satisfying basic protection demands.[29]

a. Many governments have subsidized or built transportation networks for direct protection goals (such as moving troops, supplies and tying together various regions of the nation's territory) and indirect protection goals (such as settling the club's territory). This holds for ancient states, modern non-democratic states, and modern democratic states alike. The Ancient Chinese built roads along and to the Great Wall,[30] as did the Romans and almost all modern nations, for moving troops, encouraging settlement, stabilizing food supplies, and other protection motivations. Similarly, canals have been have been aided and built out of protection motivation – e.g. Suez, Panama.[31]

The connections between government building of harbors, aid to the merchant marine, the moving of vital war supplies like oil, and naval defense has always existed. Prominent examples include the British Navy, U S privateers in the Revolutionary War, U S "Liberty" ships in World War Two, and Israeli oil tankers today. Government aid to and building of railroads has been heavily influenced by defense and other protection consideration; examples include the Canadian case, the Trans-Siberian railroad, and railroads in many Western and Central European countries.[32]

b. The institution of public education in Prussia in the eighteenth century (Frederick the Great), France in the nineteenth century (Napoleon), and to a lesser extent Great Britain in the nineteenth century [33] was primarily to provide effective troops for a national army. The desire to educate club members into the shared culture of the club has been prominent in countries allowing immigration such as the U S, Canada, and Israel.

c. Central banks can be regarded as agencies designed to protect contracts by providing a high quality (predictable) medium of exchange. Thus the valuable monopoly right attached to the right to issue paper money (seigniorage), which displaced high resource cost commodity money, was implicitly sold to the Bank of England and Second Bank of the United States in exchange for the promise that they would not overissue, i e, that they would provide "high quality money" (meaning money that would ensure as stable a price level as the displaced specie).[34,35] Later on, of course, government realized that by allowing the Central Bank to break its contract and over-issue it could finance at least cost – by the inflation tax – its other protective activities (making war).

d. According to our theory, macroeconomic policy, both counter-cyclical monetary and fiscal policy, could be viewed as attempts by governments, following Keynesian doctrine in the postwar period, to satisfy basic protection demands by reducing the variance of national income. However, the key assumption in our model – of voter rationality – would lead us to conclusions similar to the recent "rational expectations" literature,[36] which stresses the futility of discretionary monetary and fiscal policy in all but the very short-run. Indeed, the recent trends in several countries towards central banks' adoption of monetary rules, and the campaign for balanced budgets, would suggest that the acceptance of discretionary macro policy has only been a short-lived phenomenon.

Thirdly, we argue that the nature of protection provided will be related to the composition of club membership. Thus, we would expect that groups gaining protection will be those becoming the majority. Thus, in the nineteenth century, the majority of U S families owned non-human capital (which were vital inputs to producing their market income) because they were self-employed. Thus it is natural that the laws (and the court decisions interpreting and enforcing them) tended to protect non-human capital. In the twentieth century, with the decline of self-employment, the ownership of non-human capital ceased to be an important element in the market income of the majority who are now employees. At the same time, labor or human capital becomes the determinant of market income for the majority of families. As a result, the laws protecting property rights (income streams) come to give better protection to human capital-labor earnings – and weaker protection for non-human capital.

The most notable example of this shift is the legalization of unions and the right to strike de facto.[37] We observe the same phenomenon in the U S' conduct of military operations in World War II, with its emphasis on the minimizing of casualties. This same phenomenon is also illustrated in various consumer protection laws such as meat inspection, highway safety, anti-pollution, etc. The rights of owners of non-human capital are being reduced to increase the protection of human capital.[38]

Fourthly, government farm aid programs, represent very important forms of protection in high income democratic countries. These programs are very complex in their details and differ widely between countries. The protection of farm sector incomes would appear to be an example of protection of permanent income but some countries also view food as a vital war supply.

An examination of Johnson's studies of the U S farm program (1973, 1974) reveals all the key elements our theory would predict for a program of protection of a sector's permanent income: first, it has been modified and the cost reduced in response to political pressure; secondly, a key goal and achievement has been the reduction of the variance of income; thirdly, the program has not, *in practice*, been to aid low income farmers but rather farmers who are able to adjust to market forces; fourthly, the programs has included import restrictions and the subsidization of exports; fifthly, the program, for all its size and complexity, has apparently not prevented the long-run market dy-

namic of a decreasing labor input in agriculture; sixthly, US farm programs in general – including the agricultural research and extension elements – has helped preserve a sector of small business – family farms – with a rate of productivity advance exceeding the average for the U S economy.[39]

Finally, if we extend our analysis to the postwar international economy we observe that post World War Two military technology – nuclear missiles – has left the western European national clubs unable to protect themselves from external threats. The result has been the formation of a super national protection club – NATO. As we would expect, the emerging internationalization of basic protection supply has been accompanied by unprecedented levels of economic aid – the Marshall Plan, etc, – and significant reductions in barriers to trade – the European Common Market, Kennedy and Tokyo rounds, etc, between the many members of the North Atlantic protection organization.

In sum, a very wide range of major historical events, institutional developments, and changes in property rights structures can be understood in the terms of our model as responses to the demand for protection (as it existed or in its changing forms). Thus, government behavior can be explained directly from the fundamental nature of government as a protection agency without the need to assume that the governmental apparatus is *used* for additional purposes beyond its fundamental function.

## Conclusions

In this paper we have argued that the emergence of most government policies in democratic states stems from basic protection demands of the electorate. Much of the paper is then devoted to a discussion of the nature and implications of such demands. Furthermore, we argue that the assumption of rationality and competition in the political market place ensure that, over the long run, only those policies which satisfy the basic protection demands of the majority will survive.

Evidence in support of our theory is provided by an examination of a number of historical examples which describe the development of protection policies in modern democratic states.

Our approach has a number of interesting implications for the analysis of government's role in the economy.

First, we can relate our theory to some recent developments in the public economics literature. Our paper is basically concerned with positions of long-run equilibrium. We argue that in the long run, policies not efficiently providing voter-demanded protection will tend to be eliminated. Strands in the recent literature stressing the role of bureaucrats, agency costs, fiscal illusion, and special interests as determinants of government intervention can be regarded in our framework as part of the short-run dynamics. Our prediction that, in the long run, voter control will dominate, suggests a guideline (or constraints) to these short-run effects without denying their existence or importance.

Secondly, our theory suggests that the so-called "Laffer Curve" – that real per capita income growth will initially be positively correlated with the share of government expenditure in national income and then negatively correlated – is reversible. In other words, that increases in the share of total government expenditure in national income will reduce the growth rate of real per capita income, but ultimately the decline in the growth rate will induce rational voters to reduce government expenditures for all purposes other than to satisfy basic protection demands. Perhaps the recent conservative victories in several western countries are indications that such a reversal is already taking place.

Thirdly, much of the literature measuring the welfare costs of various government proposals ignores the "protection returns" attached by the public to such policies. According to our theory, so-called "inefficient" policies will persist only if the protection returns are in fact substantial.

Finally, our approach can be extended to analyze the interesting protective and representative role played by such institutions as trade unions, consumer organizations, religious organizations, etc. Thus, the apparent powerlessness of the individual in the modern democratic nation state is reduced somewhat by the existence of these organizations which act to transmit individual protection demands to government at least cost. They should not be viewed as merely representatives of special interest groups seeking transfers but primarily as brokers in the political market place.

## Footnotes

* Carleton University and University of Toronto. We would like to thank the following for helpful suggestions and comments: Carl Dahlman, Charlie Stuart, Lance Davis, Stephen Ferris, Irwin Gillespie, Bo Larsson, Robin Neill, Anna Schwartz, E G West, Stan Winer. Research was funded by a grant from Carleton University and able research assistance was provided by Paul Sevigny.

1. Unemployment compensation and other income security systems are basically insurance. However, they also give protection services. For the nation and economy as a whole they reduce the threat of social unrest and the resulting instability, thereby protecting the state and its more fortunate citizens. The danger from large numbers of dissatisfied unemployed after World War I in Great Britain was undoubtedly a factor in the significant increase in unemployment benefits there. Similarly the institution of social security in the U S was most likely related to the mass unemployment during the Great Depression.

2. The government is more efficient at supplying controlled violence because economies of scale in the production of violence exist up to and perhaps beyond the capabilities of the largest of modern states (e g, nuclear missiles).

3. See Leffler (1978) for a view of government similar to ours. The government can sub-contract the right to use violence and coercion in restricted circumstances. However, were other institutions to acquire extensive rights to use violence, then they would become de facto governments. North (1979) is also relevant here.

4. Clauses stipulating binding private arbitration of disputes between parties to a contract are sometimes termed "private enforcement". They are in fact nothing more than agreement to leave certain (difficult to formulate) parts of the contract unwritten. Enforcement of these clauses, as with the contract as a whole, is dependent on the governmental apparatus of courts, police, and prisons as the ultimate sanction against violation.

5. See Buchanan (1965) for a discussion of the club in economics.

6. See Stigler (1972) and Becker (1976).

7. An example is the movement in some countries to replace active fiscal and monetary policy as stabilization instruments with rules for monetary growth.

8. The analysis of national government policies must allow for the exogenous existence of other nation states which could attack the country in question. For a nation in a world of potentially hostile nations defense expenditures are a means of protecting national product just as insecticide sprays are a means of protecting crops against the exogenous insect threats. In both cases the exogenous threat means the protection expenditure is welfare maximizing. The fact that a world without hostile nations (or insects) could attain a higher level of welfare without expenditures on national defense (or sprays) is irrelevant.

9. The argument that government behavior is in significant elements beyond voter control is sometimes buttressed by contentions that the government is in a position to suppress systematically, and over the long-run, essential information about its activities. This argument implies that the news media – via their many and competing branches of print, radio and T V – will not find it profitable to uncover the information government wishes to suppress. Given the rewards for "investigative journalism" we find this implication unlikely.

10. In the literature there is debate if it is rational for the individual to vote when the electorate as a whole determines government policy. As Stigler (1972) points out, the determination of *issues* depends both on the size of majorities and, frequently, on more than one round of elections so that it is rational to vote even if one is only a small fraction of the majority. To this it might be added that individuals may in fact exert the strongest influence on elected officials not directly as individuals, but as members of various groups (clubs) such as churchs, unions, ethnic organizations etc. These groups can influence elected officials much more if they can "get out the vote" and thus they increase by various means the incentive for their members to vote for favored candidates.

11. See Gillespie and Labelle (1978) and Maital (1975). The findings that income redistribution has been small are not hard to explain. If voters are motivated by self-interest they will not want significant income redistributed away from themselves. The small majorities which the transfer theory postulates can force significant income redistribution would tend to be very unstable. See Stigler (1972). It is sufficient for a 49 % minority suffering redistribution to bribe 2 % of a 51 % majority receiving redistribution to change the majority and minority roles and the direction of redistributive flows. The process could then be reversed again by bribing a 2 % share out of the new majority. Obviously this kind of instability could use considerable resources in transactions with no-one gaining. Therefore the rational voters would learn – after X rounds – that redistribution towards themselves was impossible in practice and they would structure the legislative process to limit attempted redistribution.

12. See Stigler (1972).

13. In general our model is consistent with the literature of the economics of crime; see Becker and Landes (1974).

14. Following the "hedonic" price technique pioneered by Griliches we can generalize the income concept to include the value to consumers from safer products and to employees from safer working conditions.

15. We have left unanswered many questions about compulsory government insurance. One of the more

important ones is, what if 90 % of the adult population would voluntarily insure themselves? Do the protection benefits of compelling the last 10 % to insure themselves through the government exceed the costs of the resulting inefficiencies? This can only be answered by detailed studies. However, if the demand for insurance is income elastic, the answer could indeed be yes. Furthermore, benefits like a committment of the club to protect the minimum consumption levels of its elderly members would be lost with purely voluntary savings for old age.

16. See Krauss "Protection and the Welfare State" *Wall Street Journal* June 29, 1979 and his book (1978). See also Krashinsky (1978).

17. Prestige and status generating government expenditures are often considered by economists to be one of the purer forms of government "waste". However, when we study the role of military strength in protecting the club this conclusion becomes suspect. If prestige – type expenditures – e.g. a beautiful capital city or Olympic gold medals – increase the loyalty and identification of soldiers and civilian with the club they aid protect.

18. This is not to say that national clubs must have homogenous populations. Indeed, countries such as Belgium, Switzerland and Canada may be viewed as federations of homogeneous sub-clubs organized for common protection against external threats. Thus, we would expect that, as external threats diminish in importance, separatist-type movement would become a problem, e.g., the current separatist movements in Spain, France, Belgium, the U.K., Canada, etc. For an interesting study of the implications of club member homopgeneity see Landa (1979).

19. The conclusion could be much stronger if we measure the share of protection expenditure in the "net" expenditure of the federal government. By net expenditure we mean total government outlays minus the value on the market of government services of the type which could be purchased from private sector firms.

20. The Hudson's Bay and East India Companies can be regarded as situations where the British government sub-contracted its protection power to private individuals, i.e., the crown sold the monopoly right to trade and exploited vast tracts of land in exchange for an implicit promise to protect this territory from encroachment by agents of other clubs. This sub-contracting probably reflects the weakness of the British state in the seventeenth century. Indeed, in the last half of the eighteenth century much of the protection services supplied by these companies were supplemented by the use of British troops as both the cost and the complexity of protection increased because of the threat of foreign (French) encroachment. See Easterbrook and Aitken (1958) on the early role of the Hudson's Bay Company as a protection agency.

21. See Dickerson (1968) who provides comprehensive evidence that before 1763 there was little serious opposition by the colonists to the de facto operations of the Acts. Also see Egnal and Ernst (1972) and Gipson (1950) for evidence that the colonists were basically content with their position in the pre-1763 empire. Finally, see the extensive literature pointing to a small economic burden of the Navigation Acts. Harper (1939), Thomas (1965), McClelland (1969) and Reid (1970).

22. "In accounting for the radical change in attitude of many leading colonials between the years 1754 and 1774 respecting the nature of the constitution of the empire, surely among the factors that must be weighed was the truly overwhelming victory achieved in the Great War for the Empire. This victory not only freed colonials for the first time in the history of the English-speaking people in the New World from dread of the French, their Indian allies, and the Spaniards, but, what is of equal significance, opened up to them the prospect, if given freedom of action, of a vast growth of power and wealth with an amazing westward expansion. *Indeed, it is abundantly clear that a continued subordination of the colonies to the government of Great Britain was no longer considered an asset in the eyes of many Americans by 1774, as it had been so judged by them to be in 1754, but rather an onerous liability.*" (emphasis ours) Gipson (1950) in Wahlke (ed) (1962). For an alternative view see Reid (1978).

23. It took the British a full century to realize they could not properly operate an international club by direct control from London. One could hypothesize that had the British been willing to offer Dominion status to the thirteen colonies in 1776 that the American Revolution never would have occurred.

24. See Creighton (1967).

25. See Easterbrook and Aitken (1958).

26. See Mercer (1973), Neill (1979) and Wogin (1979).

27. See Easterbrook and Aitken (1958) for the standard view. Also see Dales (1967) for a neoclassical analysis of the efficiency costs of the National Policy.

28. See Innis in Neill (1972).

29. We believe that further careful historical research would allow us to explain most persistent interventions as attempts to satisfy protection demands.

30. See Reeschauer and Fairbank (1958).

31. A classic statement of the protection motivation for transportation aid was made by Gallatin, Secretary of the Treasury, to Jeffersson: "The early and efficient aid of the Federal Government is recommended by still more important considerations. The inconveniences, complaints, and perhaps danger, which may result from a vast extent of territory, can no otherwise be radically removed or prevented than by opening speedy and easy communications through all its parts. Good roads and canals will shorten distances, facilitate commercial and personal intercourse, and unite, by a still more intimate community of interest, the most remote quarters of the United

States. No other single operation, within the power of Government, can more effectually tend to strengthen and perpetuate that Union which secures external independence, domestic peace, and internal liberty." From Goodrich (1967).

32. See Heaton (1948).

33. See West (1967) and (1970).

34. See Klein (1978).

35. Peel's Act in 1844 was an explicit attempt to ensure that paper money had the same quality as specie.

36. See Sargent and Wallace (1976).

37. See Davis, Easterlin and Parker (1972) ch 6, especially pp 197–200, pp 225–227, Rees (1962) ch 1 and p 201, and Ashenfelter (1969). It is worthwhile to quote Rees at length on the role of unions in the U S because it is a classic statement of the importance of the national club and protection in economic policy: "If the union is viewed solely in terms of its effect on the economy, it must in my opinion be considered an obstacle to the optimum performance of our economic system... Many of my fellow economists would stop at this point and conclude that unions are harmful and that their power should be curbed. I do not agree that one can judge the value of a complex institution from so narrow a point of view. Other aspects of unions must also be considered. The protection against the abuse of managerial authority given by seniority systems and grievance procedures seems to me to be a union accomplishment of the greatest importance. So too is the organized representation in public affairs given the worker by the political activities of unions... If the job rights won for workers by unions are not conceded by the rest of society simply because they are just, they should be conceded because they help to protect the minimum consensus that keeps our society stable. In my judgement the economic losses imposed by unions are not too high a price to pay for the successful performance of this role." Rees (1972) pp 194–5.

38. The agitation by the media and the Eastern establishment for such laws may reflect the above average level of human capital that people in such groups have. Here it is important to ask two questions: (1) have extreme laws been on the books very long; (2) are extreme laws actually enforced? Non-enforcement may be a cheaper solution for the majority with less taste for this form of protection than getting the laws repealed.

39. The reader should not misunderstand our references to Johnson. He believes that U S farm programs which he studied either had become in the early 1970s (or always were) incorrect in major details (on equity and/or efficiency grounds). We are not suggesting otherwise! Rather, we regard the U S farm program as remarkably consistent with our model. Given the extensive criticism of these programs, we were surprised by this finding.

## References

Alchian, A and Demsetz, "Production, Information Costs and Economic Organization", American Economic Review, LXII, Dec 1972.

Ashenfelter, O and J Pencavel, "American Trade Union Growth: 1900–1960", Quarterly Journal of Economics, 83 (1969), pp 434–448.

Becker, G, The Economic Approach to Human Behavior (1976).

Becker, G and W Landes, eds, Essays in the Economics of Crime and Punishment, NBER, New York, 1974.

Becker, G and G Stigler, "Law Enforcement, Malfeasance, and Compensation of Enforcers", in R Selden (ed) Capitalism and Freedom: Problems and Prospects, Univ of Virginia, Charlottesville (1975).

Borcherding, T, Budgets and Bureaucrats, Duke Univ Press Durham (1978).

Buchanan, J, "The Political Economy of Franchise in the Welfare State", in R Selden (ed) Capitalism and Freedom: Problems and Prospects, Univ of Virginia, Charlottesville (1975).

– "An Economic Theory of Clubs", Economica, 32 (Feb 1965) pp 1–14.

Creighton, D G, "Economic Nationalism and Confederation", in Ramsay Cook (ed) Confederation, University of Toronto Press, Toronto, 1967.

Dales, J H, The Protective Tariff in Canada's Development, University of Toronto Press, Toronto, 1966.

Davis, L, R Easterlin and W Parker, eds, American Economic Growth, Harper & Row, New York (1972).

Dickerson, O M, The Navigation Acts and the American Revolution, Barnes, New York, 1863.

Easterbrook, W and H Aitken, Canadian Economic History, Macmillan, Toronto (1956).

Egnal, M and J Ernst, "An Economic Interpretation of the American Revolution," William and Mary Quarterly, Jan 1972.

Fabricant, S, The Trend of Government Activity in the United States since 1900, NBER, New York, 1952.

Friedman, D, The Machinery of Freedom, 1972.

Gillespie, W I and Labelle, J B, "A Pro-Poor or Pro-Rich Redistribution of Income," National Tax Journal, 30 (1978).

Friedman, M, A Theory of the Consumption Function, NBER, New York, 1957.

Gipson, L, "The American Revolution as an Aftermatch of the Great War for the Empire 1754–1773," in J Wahlke (ed) The Causes of the American Revolution, Lexington, 1962.

Goodrich, C, The Government and the Economy: 1783–1861, Bobbs-Merrill, New York, 1967.

Harper, L A, "The Effect of the Navigation Acts on the Thirteen Colonies," in R B Morris, ed, The Era of the American Revolution, New York, 1959.

Heaton, H, Economic History of Europe, Harper & Row, New York (1948).

Johnson, D, Gale and J A Schmittker, U S Agriculture in a World Context, Praeger, New York, 1974.

Johnson, D Gale, *Farm Commodity Programs*, AEI, Washington, 1973.

Johnson, H G, "A Theoretical Model of Economic Nationalism in New and Developing States," *Political Science Quarterly* (1965) pp 169–85.

Klein, B, "Competing Monies, European Monetary Union and the Dollar: The Case Against the Europa," *UCLA Working Paper* No 115, April 1978.

Krashinsky, M, "The Cost of Day Care in Public Programs," *National Tax Journal* 31 (1978).

Krause, M, *The New Protectionism*, International Center for Economic Policy, New York, 1978.

Landa, J T, "The Economics of the Ethnicly Homogenous Middlemen Group," *U of Toronto Inst for Policy Analysis Working Paper Series*, N 7924, Oct 1979.

Leffler, K, "Government Output and National Income Estimates," in *Carnegie-Rochester Conference Series on Public Policy*, Vol 9, *Public Policies in Open Economies* (1978).

Maitals, S, "Apportionment of Public Goods Benefits to Individuals," *Public Finance*, 30 (1975).

McClelland, P D, "The Cost to America of British Imperial Policy," *American Economic Review*, May, 1969.

Mercer, Lloyd, J, "Rates of Return and Government Subsidization of the Canadian Pacific Railway: An Alternative View, *Canadian Journal of Economics*, 6 (Aug 1973) pp 428–37.

Neill, R, "Tweedle-Dum and Fiddle-Dee-Dee," Conservatives, Liberals and the National Policy," *Carleton Economic Paper* (1979).

– *A New Theory of Value: The Canadian Economics of H A Innis*, University of Toronto Press, Toronto, 1972.

Niskanen, W, "The Pathology of Politics," in R Selden (ed) *Capitalism and Freedom: Problems and Prospects*, U of Virginia, Charlottesville (1975).

Nordhaus, W and J Tobin, "Is Economic Growth Obsolete?" in *Economic Growth*, NBER, Fifth Anniversary Colloquium V. (New York 1972).

North, D C, "A Framework for Analysing the State in Economic History," *Explorations in Economic History*, 16 (1979).

Rees, A, *The Economics of Trade Unions*, U of Chicago Press, Chicago, 1962.

Reeschauer, X and X Fairbank, *East Asia: The Great Tradition*, New York, 1958.

Reid, J, "Economic Burden: Spark to the American Revolution," *Journal of Economic History*, Vol XXXVIII, No 1, March, 1978.

Reid, J, "On Navigating the Navigation Acts with Peter D McClelland: Comment," *American Economic Review*, Dec 1970.

Sargent, T and N Wallace, "Rational Expectations and the Theory of Economic Policy," *Journal of Monetary Economics*, II, April 1976.

Schumpeter, J, *Capitalism, Socialism and Democracy*, 3d ed, New York, 1950.

Stigler, G, "The Economists' Traditional Theory of the Economic Functions of the State," Ch 7 in *The Citizen and the State: Essays on Regulation*, University of Chicago (1975).

– Economic Competition and Political Competition," *Public Choice*, 13 (Fall 1972).

– "General Economic Conditions and National Elections," *American Economic Review*, May, 1973.

Thomas, R P, "A Quantitative Approach to the Study of the Effects of British Imperial Policy Upon Colonial Welfare: Some Preliminary Findings, *Journal of Economic History*, Dec 1965.

Tullock, G, *The Social Dilemma: The Economics of War and Revolution*, University Publication, Blacksburg, Va, 1974.

Wahlke, J (ed), *The Causes of the American Revolution*, revised ed, D C Heath, Lexington, 1962.

West, E G, *Education and the State*, 1970.

– "The Political Economy of American Public School Legislation," *Journal of Law and Economics*, Oct 1967.

Wogin, G, "Land Settlement and Freight Rate Policies of the CPR," unpublished, 1979.

# Illusions in Fiscal Policy: A Case Study

## BY FRIEDRICH SCHNEIDER AND WERNER W POMMEREHNE*

*The paper examines government fiscal policy in a representative democracy under the conditions of fiscal illusion, i.e. the systematic misperception by voters of the public revenue burden they bear and the benefits they receive through government's spending policy. Empirical evidence is offered for Australia. The results indicate that voters seem to be rationally underinformed and, thus, subject to systematic misperception of fiscal variables and that government tries to exploit such fiscal illusions when working towards its re-election.*

## I. Information, Rational Ignorance and Fiscal Illusion

Traditional consumer theory implicitly assumes perfect information or, at least, a very short learning span. Some economists have modified this assumption and explicitly analyzed the cost of information to the consumer, thereby recognizing what may be called *rational ignorance* in consumer choice.[1] Due to the positive and increasing marginal costs of collecting information, consumers may not have incentives to acquire complete information.

There are differences in the costs of acquiring information which reflect differences in the nature of the goods involved. Some goods such as cars have the characteristics of "search" goods, i.e., it pays to invest in an information search *prior* to purchase to obtain a better idea of quality. Other goods, like brands of red wine or cheese, are "experience" goods for which information obtained through prior search is less appropriate than information derived from purchase and experience. For these goods quality becomes known after purchase.[2] There are also certain goods whose qualities are difficult to judge even after purchase, i.e., they have "credence" qualities and expertise is required in their evaluation. (Darby and Karni [1973]). Credence qualities are characteristic of goods and services that are utilized in combination with other goods and services composed of uncertain properties. However, it should be noted that emphasis is usually placed on the *complexity of the good or service itself* as the source of rational ignorance. The property rights in such a case may,

in principle, be well established permitting the existence of a competitive market.

Rational ignorance is not restricted to the marketplace, however, as was recognized very early by Anthony Downs (1957) and other economists in their analyses of the voting process under majority rule. Their basic argument is that the influence exerted by a single individual through his vote is insignificant in determining the outcome of a national election because the probability of his being the decisive vote may very quickly approach zero and he is, therefore, likely to invest little or no effort in obtaining information on public policy. It follows that he is rationally ignorant.[3] There are also positive information costs arising from the *nature* of publicly provided goods and services and their financing which are often much higher than in the case of marketed goods. Evaluation of publicly provided goods and services may require a greater degree of expertise. The main difference between the political and the market sphere lies, however, in the significance of political *institutions*. In the political sphere, property rights are *not* well defined in terms of the goods and services publicly provided. There may be major differences between political parties, but the common principle is that only one majority can exist at a time and, therefore, the winning political party has a *legal monopoly* over time.

Fiscal illusion can be seen as a special kind of rational ignorance. It is defined here as the *systematic* misperception by people of the fiscal burden they bear and the benefits they receive through public policy. It is due principally to the spe-

cific nature of the various public revenue items and public services and the associated information costs. Moreover, it is reinforced by the individual's limited influence on the political process and the weak incentives to acquire information about the policies under discussion. It should be noted that the term "illusion" is *not* intended to convey the idea that individuals are somehow behaving irrationally. The systematic misperception of the burden and benefits arises because it is not in the individual's narrow self-interest to invest in accurate information and knowledge about the public sector. Despite its somewhat unfortunate connotations, "illusion" is a term that has become widely used and, thus, will also be used here.

In this paper fiscal illusion will be analyzed by assuming that we have (i) a government which aims to maximize its utility subject to various constraints and (ii) an electorate that is also utility maximizing.[4] We will show in Part II that voters are rationally underinformed and, thus, subject to misperception of fiscal variables, and that when this is the case governments will try to exploit the situation in order to reach their ideological goals or to improve their re-election chances.

The first question considered is whether, taking into account the explicit preferences of the governing party for specific goals and instruments, the government will try to exploit fiscal illusions. If so, will this result in a predictable pattern of changes in the *structure* of the fiscal variables over a legislative period. This is examined in Part III in an empirical analysis of the behavior of the Australian federal government using monthly data for the period 1970–78.[5]

The second question is whether the economy is manipulated *cyclically* over the legislative period in such a way that the government makes a favorable impression on the voters, mainly to improve its re-election chances. This issue is examined in Part IV by extending the analysis back to 1960, using quarterly data, in order to cover a larger number of legislative periods. Part V presents a summary and a discussion of the general approach.

## II. Voters' Behavior in a Representative Democracy

It is assumed that in evaluating the government's performance voters maximize their own utility, i.e., the degree of their support, as measured, for example, by current government popularity or by voting for the government at election time, de-

pends on how satisfied they are with its overall performance. Because of the high costs and low benefits involved for the individual there is little, or even no, incentive to become fully informed about overall performance. This is true with regard to obtaining information not only about one's present and future burden of government receipts and benefits received through public expenditures, but also about the past, which may be discounted by the voters.

### 1. Illusions about Taxing and Spending

As discussed above, individuals will generally have little incentive to become fully informed about the government sector. Besides, one can expect that the information absorbed on a day-to-day basis may be biased in a systematic way because of the different costs involved in acquiring information on different revenue items. These information costs are dependent on the varying degrees of visibility, the timing of the extraction items and the degree of hidden shifting of the revenue burden.[6] Thus, it appears to make little sense for the individual to secure complete information on the different parts of the fiscal burden borne since the marginal benefits of doing so quickly approach zero. From this we may conclude that the voter/taxpayer perceives certain public revenue items less comprehensively than others or not at all and that this opens up opportunities for the government to install revenue extraction institutions that will decrease the perceived cost of government, *ceteris paribus,* thereby favorably influencing the voters' evaluation of the government's performance.

Voters may also be subjected to systematic misperception of the benefits of public spending programs. However, it is very difficult to derive a hypothesis about systematic biases in individuals' perceptions of public services. This problem arises because of the difficulty in discriminating between the actual and the perceived benefits derived from publicly provided goods. The cost of the government's activities to a taxpayer via taxes and other forms of fiscal extraction is, to a large extent, identifiable so that if the perception of this cost can be reasonably determined the difference can be attributed to fiscal illusion. This does not hold for the spending side, however. For this reason we will not try to distinguish here between preferences and favorably biased perceptions when dealing with the government's spending.

We also assume that, as is shown in various

studies,[7] voters rarely draw a close connection between public services and their actual financing sources. This assumption may be valid as the principle of non-affection is used at the federal level.[8] The link between federal revenue and expenditure is further obscured by government borrowing.

When empirically examining misperceptions we also have to consider how the government's ability to steer the overall economy is evaluated by the voters. It can be expected that the individual's impression of this will be influenced by his perception of the current and past states of the economy, mainly represented by the most visible variables such as the rate of unemployment and the rate of inflation. A satisfactory state of the economy, i.e. a low rate of unemployment and/or inflation, will help create a favorable image. Again, because the marginal benefits very quickly approach zero for the individual trying to obtain full information, it is rational to limit oneself to what can be easily culled from day-to-day sources. This, in combination with the voters' short memory, results in the current economic situation being taken as the main indicator in evaluating the government's economic performance and in past events being discounted or even forgotten.

## 2. The Empirical Search for Fiscal Illusions

There are several ways to examine whether voters are subject to systematic misperceptions.[9] For reasons relating mainly to the availability of data, we have relied on the following procedure: first, we modelled the government's ability to steer the economy, primarily its ability to fight unemployment and inflation, as this is perceived by the voters; secondly, we tried to explain the voter's perception and evaluation of overall government activity by taking an indicator of the perceived economic performance and a series of structural variables representing the government's revenue and spending sides.

A monthly Gallup survey taken since January 1970 asking "Are you satisfied (or not) with the current economic performance of the government?" was used to measure how the voters perceive and evaluate the government's ability to steer the economy. This gives us our first dependent variable. As independent variables we have used the rates of unemployment and inflation and the degree of industrial unrest (measured by the number of working days lost because of strikes), the

last factor being included in order to capture major economic disturbances for which the government is held, at least partially, responsible by the Australian electorate. The second dependent variable is current government popularity,[10] taken as an indicator of the government's overall performance.

Turning to the independent variables, we have grouped the items on the government revenue side into: two classes of direct taxes, namely, pay-as-you-earn income tax and assessed income tax, and a tax on dividends and interest; one class of indirect taxes (sales tax, company taxes and license fees); one of customs and various duties; and increases in government debt.

The various spending items are also grouped into classes of expenditures: health, education, general and scientific research, culture and recreation; public administration, law and order, public safety, and legislative services; investments in transportation and communication, water and electricity supply, other public utilities, and investment grants to the states; defense expenditure; and various transfer payments (social security and welfare payments to individuals, payments to disabled people and retraining programs). All variables are measured as a share of total government revenue or expenditure respectively. The ratio of total federal revenue to GNP is introduced to take into account perceptions of the increase in the overall federal revenue burden that occurred during the 1970s.[11]

With regard to economic performance we expect negative signs for the coefficients of unemployment, inflation and industrial unrest. In the popularity function we expect a negative sign and a significant coefficient only for those revenue shares which are strongly perceived and a positive sign for those expenditures which are favorably perceived by a large part of the voters. All data are monthly and a three-month lag is taken for all independent variables as it is assumed that the citizen needs some time to notice a change in taxation and spending as well as in the economic situation.[12] With respect to the voter's memory, we assume for simplicity that this can be stated by lagged endogenous variables and that the same discount rate holds for all the independent factor in each equation, i.e., a Koyck transformation is used. The results of the simultaneous GLS (general least square) estimation of both equations for the overall period 1971:1 through 1977:9 are given in the following two equations:

(1) $GEP_t = 12.47 + 0.48^{**}GEP_{t-1} - 1.65U_{t-3}^{**}$
   $\quad\quad\quad\quad (10.56) \quad\quad (-4.45)$
   $\quad\quad\quad\quad\quad\quad\quad\quad\quad [-0.64]$

   $\quad\quad - 0.18^{**} I_{t-3} - 0.39^{*} InUn_{t-3}$
   $\quad\quad (-2.76) \quad\quad\quad (-2.03)$
   $\quad\quad [-0.09] \quad\quad\quad (-0.16)$

   d.f. = 77, $\bar{R}^2$ = 0.88, h = 1.36

   List of symbols:

   GEP = government's economic performance
   U = rate of unemployment
   I = inflation
   InUn = industrial unrest

(2) $POP_t = 0.43 + 0.48^{**}POP_{t-1} + 0.37^{**}GEP_t - 1.81^{**}TotRev_{t-3}$ -
   $\quad\quad\quad\quad (9.36) \quad\quad\quad (5.41) \quad\quad (-6.80)$
   $\quad\quad\quad\quad\quad\quad\quad\quad\quad [0.57] \quad\quad (-0.46)$

   $\quad\quad - 0.81^{**}TaxEaIn_{t-3} \quad - 0.36^{*}TaxInDiIn_{t-3} \quad + 0.21\ IndTax_{t-3} \quad + 0.82GovDept_{t-3}$
   $\quad\quad (-6.54) \quad\quad\quad\quad (-2.61) \quad\quad\quad\quad (1.30) \quad\quad\quad (1.80)$
   $\quad\quad [-1.21] \quad\quad\quad\quad [-0.47] \quad\quad\quad\quad (0.27) \quad\quad\quad [0.21]$

   $\quad\quad + 0.44^{**}TranPay_{t-3} \quad + 0.14^{*}ExpHeaEdRec_{t-3} \quad + 0.03\ InvTransWaElectr_{t-3}$
   $\quad\quad (4.43) \quad\quad\quad\quad (2.23) \quad\quad\quad\quad\quad (1.09)$
   $\quad\quad [0.68] \quad\quad\quad\quad [0.13] \quad\quad\quad\quad\quad [0.07]$

   $\quad\quad - 0.48\ ExpPubAdmLawPuSaf_{t-3}$
   $\quad\quad (-1.30)$
   $\quad\quad [-0.09]$

   d.f. = 63, $\bar{R}^2$ = 0.96, h = 1.07

   List of symbols:

   POP = government's popularity
   GEP = government's performance
   TotRev = total revenue as a share of GNP
   TaxEaIn = tax on earned income as a share of total revenue
   TaxInDiIn = tax on income from dividends and interest as a share of total revenue
   IndTax = indirect taxes as a share of total revenue
   GovDebt = government debt incurred as a share of total revenue
   TranPay = transfer payments as a share of total expenditure
   ExpHeaEdRe = expenditure on health, education and recreation as a share of total expenditure
   InvTransWaEl = investment in transportation, water supply and electricity as a share of total expenditure
   ExpPubAdmLawPuSaf= expenditure on public administration, law and order, and public safety as a share of total expenditure.

The figures in parentheses below the parameter estimates indicate the t-values; one asterisk indicates statistical significance at the 95 % confidence level and two asterisks refer to the 99 % confidence level, using a two-tailed test. The figures in brackets are the elasticities for the variables (estimated under a double logarithmic specification of equations 1 and 2); d f shows the degree of freedom; $\bar{R}^2$ is the corrected coefficient of determination; and h indicates the Durbin test statistics for autocorrelation.

The results show that citizens do, indeed, discount economic events highly (eqns 1 and 2) as well as changes in the revenue and spending structures (eqn 2). The coefficients of the lagged endogenous dependent variables indicate that over 95 % of what happens is forgotten within the space of one year. This is an important factor in the voters' evaluation of the state of the economy and in their evaluation of the changes in the various revenue shares (as proxies for the actual fiscal burden) and in the spending shares (as proxies for preferences in and perceptions of expenditure benefits).

The results for equation 1 show that all three indicators have a significant negative impact on government economic performance, with the rate of unemployment by far the most significant. If the rate of unemployment rises by one percentage point, perceived government economic performance drops, ceteris paribus, by 1.65 percentage points. The same general results show up when a quarterly model is used and the time period is extended back to the late 1950s (see Schneider and Pommerehne, [1980]).

Equation 2 shows that the total revenue burden as well as two revenue and two spending shares have a strongly significant impact on the government's current popularity in addition to the highly significant influence of perceived economic performance. Surprisingly, the marginal influence of the variable for government's economic performance is quite small: a decrease of one percentage point in the perceived economic performance lowers current popularity, ceteris paribus, by only 0.37 percentage points. In contrast, the marginal impact of the total revenue burden is almost five times as large.[14] The individual revenue items exhibit a highly significant negative influence on government's current popularity only in the case of direct taxes. The respective elasticities are −1.21 for direct tax on earned income, and −0.47 for direct tax on income from dividends and interest. This result is in line with the classic argument in the literature on public finance that direct taxes are most strongly felt whereas indirect taxes and most other revenue items, especially government debt, are much harder to detect.[15] Looking at the spending side, we find a significant positive marginal impact on government popularity only in the cases of transfer payments, which go mostly to private households, and of expenditure on health and education programs. The impact of the latter, however, is only one-third of the first. The relatively high marginal influence of transfer payments is not implausible as it may be argued that they strongly reduce uncertainty concerning future direct benefits and, thus, may be valued highly by voters. Even if not all voters/taxpayers will end up being public transfer beneficiaries, the great majority of them may still expect to do so. Comparing the influences of both sides of the government budget, the total marginal influence of the revenue items is five times as large in absolute terms as that of the spending items.[16]

In order to check our empirical results we made an ex ante forecast for the period 1977:10 through 1978:12 (15 observations), based on the GLS estimates for both equations over the period 1971:1 through 1977:9. The results are shown in Table 1.

Table 1: Ex ante forecast for Australian government economic performance (GEP) and government popularity (POP); 1977:10 to 1978:12[a]

| Statistical measures | Government economic performance (GEP) | Government popularity (POP) |
|---|---|---|
| Average root mean squared-error | 1.07 | 0.62 |
| Theil's inequality coefficient | 0.11 | 0.05 |
| Average mean error of deviation (in %) | 2.10 | 0.63 |

[a] The theoretical values of the lagged endogenous variable and the actual for the exogenous variables were used for the calculations.

Table 1 shows that both ex ante predictions are clearly superior to naive forecast methods, (Theil's inequality coefficients being much smaller than 1), indicating that the most important variables determining the perceived government's economic performance and its popularity have been taken into account. For the popularity forecast the average root mean squared error is less than one percentage point, i.e., the predicted development comes very close to the actual outcome.

In summary, therefore, our empirical results show that changes in *certain* types of taxes and other public revenues are indeed harder for voters to detect than others and, therefore, a government which relies on hard-to-detect revenues can, *ceteris paribus*, expect to be more popular. Similarly on the spending side, additional transfer expenditure and, to a lesser extent, additional expenditure for certain services such as health and education seem to be more popular and more strongly felt than increases in other expenditure items. These effects on both sides of the budget and the fact that voters discount highly past government activities provide opportunities for the government to use its fiscal instruments in a systematic way to retain the popularity it needs to allow it to either achieve its ideological goals or to secure re-election.

## III. Government's Policy I: The Strategic Use of Fiscal Structures

As shown in the previous section, there are possibilities for the government to take advantage of voters' misperceptions through the systematic use of fiscal policy instruments. The extent to which such strategic use will be made is heavily dependent on the constraints placed on the government.[17] Even if we assume that a government cannot be voted out of office in the middle of a legislative period it is still subject to various constraints in trying to achieve its ideological goals, of which the most important is the re-election constraint. Thus, the government faces a dynamic maximization problem of determining when to undertake whatever fiscal policy action is required to maximize its utility.

### 1. Government Behavior

We assume that the government regards the result of popularity surveys as the best current indicator of its re-election chances. If its current popularity is high and/or if there is plenty of time left until

the next election, the government will use its various fiscal instruments to pursue its ideological goals.[18] In comparison with a right-wing government, a left-wing government will generally increase public sector activity by introducing new programs and/or expanding existing programs. Looking at Australian governments in the 1970s, the Labor (left-wing) government explicitly stated preferences for more spending on education, improvements in welfare and health care systems and decreased outlays for national defense. The Country-Liberal (right-wing) governments stated preferences for a much smaller level of growth in current and future government activities and a strengthening of private sector activities. There were also major differences in preferences on the methods of financing public expenditures. Whereas a Country-Liberal government tended to favor tax financing, a Labor government relied more on incurring additional debt.

If re-election chances are indicated to be poor by a low popularity standing and/or the next election is close, the government will concentrate on securing re-election rather than on pursuing its ideological goals, which it can in any case only hope to put into effect by remaining in power. For this purpose, the government, regardless of who is in power, will try to create favorable fiscal illusion on the part of the voters by means of a systematic revenue and spending policy, counting on the voters' short memory to aid it in this. Before an election it will:

(i)  reduce direct taxes, such as the personal income tax, and

(ii) increase harder-to-detect revenues especially public debt in order to finance additional popular spending or to cover the deficit caused by the lowering of strongly felt revenues.

We assume that in a state of low popularity, which dampens the chances of being re-elected, the government, when formulating the use of fiscal policy instruments, will react all the more strongly the greater is the ratio $POP^*/POP_t$ (the critical level of current popularity $POP^*$ has the value of 51 %). The second important factor, the amount of time since the last election, i e, the inverse of time left until the next election, is reflected in the discretionary variable TSLE (time since last election) which takes the values 1, 2, 3,...,36, starting with the beginning of each legislative period. This formulation implies that the government's reaction six months before an election will

be stronger than it is when the election is still far off. The government also has to take legal obligations and the behavior of the public administration into consideration when using its revenue and spending instruments. In most Western democracies it is realistic to assume that the public administration does not simply follow the wishes of the government but rather tries to maximize its own utility. Most members of the public administration show risk-averse behavior, tending to resist major changes because this may threaten their own position and preferring to make only small and incremental changes. As a consequence the government may be considerably handicapped in trying to carry out substantial and sudden changes in its spending and revenue policies. It would, therefore, seem to be useful to take past spending and revenue structures into consideration as these are the starting points for changes therein.

There are two economic constraints in addition to the legal and administrative ones: (1) The government has to take changes in the balance of payments into account. The importance of this outside restriction on economic policy has, of course, been noted by many other authors (e.g. Barry and Guille [1976]). (2) The size of the budget deficit (a positive sign in the case of deficit, and a negative sign for budget surplus), whose maximum equals the maximum incurrable additional debt, is set through the legal framework. To measure whether a budget deficit is really binding or not, its deviation from the long-term trend of the deficit is considered here. We assume that if the current deficit lies above this trend, the government has to consider it as a constraint; while if it is below, there is still some leeway for additional spending and/or tax cutting.

It is now possible to formulate the following equation for the i'th fiscal instrument (INST$^i$):

$$
(3) \quad INST_t^i = \left\{
\begin{array}{ll}
a_1 & INST_{t-12}^i \\[2ex]
+a_2 & \text{(Change in balance of payments)}_{t-6} \\[2ex]
+a_3 & \text{(Budget deficit deviation from the long-term trend)}_{t-6}
\end{array}
\right.
$$

Administrative and economic constraints

$$
\left\{
\begin{array}{ll}
+a_4 & \text{(Current popularity standing)}_{t-6} \\[2ex]
+a_5 & \text{(Time since last election)}_t
\end{array}
\right.
$$

Political constraints

$$
\left\{
\begin{array}{ll}
+a_6 & \text{(Ideological preferences of Country-Liberal governments)}_t \\[2ex]
+a_7 & \text{(Ideological preferences of Labor Governments)}_t
\end{array}
\right.
$$

Ideological preferences

for $i = 1, 2, \ldots, 26$[19]

The 26 instruments refer to 10 federal revenue and 16 spending items and are calculated as shares of total revenue and total expenditure respectively in order to capture the changes in the structure and eliminate the typical time trend effects. The explanatory variables are lagged by six months as we assume that the Australian government needs at least half a year to react to political and economic changes.[20]

The theoretically expected signs of the administrative and economic constraints are $a_1 > 0$ for both revenue and spending sides, $a_2 < 0$ and $a_3 > 0$ for revenue items and $a_2 > 0$ and $a_3 < 0$ for expenditure items. In the cases of $a_4$ and $a_5$ we expect a decrease in share for those revenue items which are strongly felt as a burden by the voters and an increase for spending items which are popular. For the two ideological parameters $a_6$ and $a_7$ we

Table 3: Policy functions of the Australian government for the spending side under the assumption of pursuing ideological goals, GLS-estimates, 1971:1 to 1977:9, ex ante forecast 1977:10 to 1978:12[a]

| Instruments on the spending side (as share of total expenditure) | Legal, administrative, and economic constraints | | | | Re-election constraints | | Ideological preferences | | Test statistics | | Ex ante forecast | | Mean error (percent deviation) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Lagged endogenous instrument | Change of balance of payments | Budget deficit deviated from long term trend | Payment of interest on govern debt | Current popularity standing | Time since last election | Country /Liberal | Labor | $\bar{R}^2$ | h | Root mean squared error | Theil's inequality coefficient | |
| | (t-12) | (t-6) | (t-6) | (t-6) | (t-6) | (t) | | | | | | | |
| Social security and welfare payments to individuals | 0.91** (9.71) | -0.15 (-1.59) | -0.09* (-2.43) | -0.06* (-2.61) | 0.26** (2.89) | 0.30** (3.04) | 0.59 (1.74) | 3.16** (2.84) | 0.78 | 1.03 | 0.90 | 0.13 | 1.03 |
| Financial aid to disabled persons | 0.78** (8.49) | 0.21** (2.89) | -0.11** (-2.89) | -0.09* (-2.59) | 0.11* (2.27) | 0.15* (2.33) | 2.41* (2.36) | 1.40 (1.86) | 0.69 | 1.21 | 1.43 | 0.37 | 1.45 |
| Retraining programs | 0.96** (9.43) | -0.12 (-1.31) | -0.09* (-2.54) | -0.06* (-2.16) | 0.13* (2.54) | 0.14** (2.86) | -0.84 (-0.98) | 2.56* (2.36) | 0.72 | 1.36 | 0.69 | 0.27 | 1.31 |
| Foreign aid and overseas grants | 0.63** (3.64) | 0.09* (2.31) | 0.31 (1.59) | 0.07 (1.39) | -0.41 (-1.86) | -0.43 (-1.91) | 1.32 (0.86) | 2.49* (2.03) | 0.49 | 1.58 | 1.71 | 1.29 | 1.96 |
| Grants to the states | 0.32** (2.89) | 0.08* (2.11) | 0.36 (1.39) | 0.21 (1.39) | -0.32 (-1.58) | -0.29 (-1.89) | 2.47 (1.86) | -1.54 (-1.64) | 0.48 | 1.64 | 2.32 | 1.31 | 2.13 |
| Education | 0.86** (8.99) | 0.09 (1.86) | -0.08* (-2.08) | -0.11** -2.81 | 0.39* (2.32) | 0.36** (2.89) | 2.13 (0.97) | 8.47** (3.56) | 0.76 | 1.09 | 1.36 | 0.33 | 1.21 |
| General and scientific research | 0.83** (7.59) | -0.23 (-1.51) | -0.09* (-2.59) | -0.05* (-2.64) | -0.21 (-1.39) | -0.13 (-1.47) | -0.47 (-0.56) | 1.03 (1.84) | 0.43 | 1.77 | 1.39 | 1.21 | 2.46 |
| Culture and recreation | 0.73** (5.91) | -0.20 (-1.27) | -0.12** (-2.80) | -0.09* (-2.51) | -0.03 (-0.98) | -0.17 (-1.04) | 0.38 (0.96) | 1.42 (1.84) | 0.40 | 1.73 | 1.32 | 1.17 | 3.45 |
| Health and hospital | 0.89** (8.47) | 0.29* (2.40) | -0.21** (-2.84) | -0.13** (-2.91) | 0.20* (2.44) | 0.21** (2.93) | 0.86 (0.94) | 6.58** (3.59) | 0.72 | 1.13 | 1.29 | 0.36 | 1.36 |
| Defence | 0.91** (5.03) | 0.28** (3.21) | -0.13* (-2.69) | -0.05* (-2.51) | -0.09 (-1.36) | -0.23 (-1.84) | 10.47** (3.59) | -2.86* (-2.14) | 0.72 | 1.09 | 0.80 | 0.27 | 1.33 |
| Administrative services | 0.93** (9.36) | -0.06 (-0.94) | 0.11 (1.56) | 0.21 (1.59) | 0.09* (2.03) | 0.10* (2.14) | -0.86* (-2.14) | 3.89* (2.54) | 0.71 | 1.22 | 1.29 | 0.47 | 1.39 |
| Legislative services | 0.98** (8.51) | -0.25 (-1.27) | 0.18 (1.06) | -0.08* (-2.02) | 0.09 (1.54) | 0.08* (2.03) | -0.59 (-1.34) | 1.13* (2.03) | 0.60 | 1.49 | 1.09 | 0.51 | 1.84 |
| Law and order, public safety | 0.87** (8.35) | -0.15 (-1.12) | 0.19 (1.31) | 0.29 (1.65) | 0.13* (2.06) | 0.08* (2.32) | 1.84** (2.77) | -0.74 (-0.86) | 0.76 | 1.36 | 0.88 | 0.41 | 1.23 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Transport and communication | 0.81** (4.64) | 0.18* (2.69) | -0.21* (-2.74) | -0.14** (-2.89) | 0.06 (1.54) | 0.12 (1.81) | 7.45* (2.47) | 0.38 (0.22) | 0.68 | 1.27 1.21 | 0.36 | 1.49 |
| Water supply and electricity | 0.66** (5.51) | 0.17* (2.51) | -0.09* (-2.61) | -0.10** (-2.79) | 0.03 (1.66) | 0.13* (2.06) | 2.54* (2.57) | -0.86 (-0.39) | 0.63 | 1.44 0.82 | 0.41 | 1.58 |
| Other public utilities | 0.86** (9.27) | 0.24* (2.09) | -0.07* (-2.11) | -0.18* (-2.56) | -0.41 (-1.86) | -0.42 (-1.91) | -0.87 (-1.94) | 6.54** (2.81) | 0.51 | 1.62 0.99 | 0.94 | 1.89 |

$a$   For notes see Table 2.

As with regard to the revenue side, the administrative and economic constraints have in most cases a highly significant effect on the government's use of its spending instruments. If we look at the ideological differences between the two types of government, we see that these result in very different spending behavior when they have leeway to exercise their ideological preferences. Country-Liberal governments favor additional expenditure on capital formation (mainly transportation, communication, and other infrastructure projects), defense and law and order. A Labor government prefers to decrease the last two and instead favors additional transfer programs (social security, foreign aid) and the expansion of education, health care and the remainder of the public service sector.

If the government is afraid that it will not be re-elected, the first three of the four transfer items will be further increased in addition to the most favorable perceived expenditures, i.e., on education and health. For the remaining spending items we find no statistically significant and quantitatively important evidence of their being specifically manipulated to help secure re-election. Only conservative governments deviate greatly from their ideological instrument preferences when running for re-election. For a Labor government, the instruments used to achieve their ideological goals are more similar to those needed to secure re-election.

The *ex ante* forecasts which have been made to test the model's predictive ability in an election year lead to superior results for 11 of the 16 spending items (evaluated by Theil's inequality coefficient, which is smaller than 1). The instruments for securing re-election again give the best forecast results, with an average percentage mean error of less than 1.5 %. If we compare these results to our finding for the revenue side, we see that the use of spending instruments is of less importance in securing re-election. This is not implausible and may simply reflect the often stated rigidity of the expenditure side as compared to the revenue side.[22]

## IV. Government's Policy II: The Cyclical Use of Fiscal Instruments

Besides making systematic changes in the structure of the revenue and spending institutions, a government can use its fiscal instruments in a

cyclical way to improve the economic situation before an election. As discussed in Part II, voters are in a state of rational ignorance when evaluating the government's performance in steering the economy. This gives the government an additional possibility to influence the voters' current perception of the state of the economy. As a low rate of unemployment has the greatest value in the voters' eyes (see eqn 1)[23] the government will adopt an expansionary fiscal policy before an election. This results in an increase in government expenditures in relation to the general trend,[24] and in smaller increases or even decreases in the strongly felt revenue items. The government will, in particular, prefer to increase its public debt to finance any additional deficit that arises. It is very difficult for the opposition party (or parties) to argue convincingly against such a policy that purports to having the fight against unemployment as its main goal.

If the government succeeds in stimulating the economy before an election, the automatic increase in the tax revenues that follows assures additional financial assets at a later point in time. Direct taxes will automatically increase because of the resulting real or inflationary economic growth or they may be discretionarily increased by the government after the election with the argument that this must be done to balance the spending and revenue accounts and/or to fight inflation.[25] Again the opposition will find it difficult to counter this policy convincingly. Even more important, the voters will forget these unpopular measures if the measures are manipulated properly and reversed as the next election approaches. This removes an incentive for the government not to act this way.

In order to test the hypothesis that the government will adopt an expansionary policy to create a favorable economic situation before an election a longer time period was considered: 1960:II through 1976:IV, using quarterly data. The growth rates for the fiscal instruments were considered to take into account the general trend of government fiscal activity over the last two decades,[26] with the equations for the use of revenue and spending instruments again being estimated simultaneously. The results are given in Table 4 which also includes the *ex ante* predictions for 1977:I through 1978:IV.

Table 4: Policy functions of the Austdralian government under the assumption of pursuing ideological goals; GLS-estimates, 1960:II to 1976:IV, ex ante forecast 1977:I to 1978:IVa

| Instruments on the revenue and spending side (growth rates) | Legal, administrative, and economic constraints | | | Re-election constraints | | Ideological preferences | | Test statistics | | Ex ante forecast | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Lagged endogenous instrument | Change of balance of payments | Percent of budget deficit deviated from long term trend | Current popularity standing | Time since last election | Country-/Liberal | Labor | $\bar{R}^2$ | h | Root mean squared error | Theil's inequality coefficient | Mean error (percent deviation) |
| | (t−4) | t−2) | (t−2) | (t−2) | (t) | | | | | | | |
| Tax on earned income | 0.36** (7.54) | −0.44** (−3.74) | 0.36** (3.94) | −0.59** (−7.08) | −0.89** (−7.56) | 5.49** (3.56) | 2.49** (2.79) | 0.84 | 1.17 | 1.92 | 0.74 | 1.32 |
| Tax on income from dividends & interest | 2.53** (5.68) | −0.57** (−4.22) | 0.42** (3.12) | −0.27** (−3.58) | −0.61** (−4.12) | 2.47* (2.57) | 3.56* (2.56) | 0.79 | 1.21 | 2.21 | 0.84 | 2.51 |
| Indirect taxes | 2.03** (6.89) | −0.36** (−3.99) | 0.45** (3.56) | −0.19 (−1.21) | 0.18 (0.59) | 2.89* (2.21) | 4.59* (2.56) | 0.65 | 1.36 | 2.59 | 0.94 | 2.31 |
| Government debt incurred | 0.89** (4.07) | −0.18* (−2.21) | 0.21* (2.36) | 0.34 (0.98) | 0.44 (1.08) | 2.59* (2.36) | 4.56** (2.88) | 0.59 | 1.51 | 2.84 | 1.31 | 2.39 |
| Transfer payments | 4.79** (8.54) | 0.17 (1.98) | −0.31** (−3.54) | 0.38** (3.55) | 0.47** (4.59) | 2.57** (2.72) | 4.50** (2.99) | 0.81 | 1.13 | 1.87 | 0.80 | 1.27 |
| Expenditure for health, education and recreation | 3.59** (7.59) | 0.37** (3.56) | −0.27** (−3.09) | 0.27** (3.61) | 0.49** (3.58) | 3.59* (2.54) | 3.84* (2.69) | 0.82 | 1.21 | 1.80 | 0.83 | 1.35 |
| Investment in transportation, water supply, electricity | 0.75** (3.59) | 0.49** (4.11) | −0.33** (−3.59) | 0.19 (1.74) | 0.21 (0.84) | 2.57* (2.17) | 0.89 (1.13) | 0.74 | 1.33 | 2.13 | 0.91 | 2.13 |
| Expenditure for public administration, law & order, public safety | 1.21** (4.69) | 0.22** (2.79) | −0.27* (−2.54) | 0.08 (0.77) | 0.06 (0.99) | 3.40** (2.89) | 3.99** (3.12) | 0.72 | 1.41 | 2.54 | 0.89 | 2.21 |

a  For notes see Table 2.

The results confirm our hypothesis on the cyclical use of fiscal instruments over a legislative period.[27] When trying to secure re-election, direct taxes are significantly reduced or at least not discretionarily increased and transfer payments, mostly to private households, are increased as expected. On the other hand, indirect taxes and public debt are used to finance the additional expenditures aimed at stimulating the economy. *After* an election the opposite use of fiscal instruments can be observed. Our additional check of these empirical results, the ex ante forecast for the period 1977:I through 1978:IV, indicates that, again with the exception of government debt, the predictions are far superior to naive forecasts.

## V. Summary

Voters' illusions concerning the fiscal burden and benefits of public expenditures have been considered as a consequence of their reluctance to undertake a costly information search. This results in a discrepancy between the perceived and actual consequences of particular fiscal institutions. Our general hypothesis is that indirect and hidden revenues will result in individuals being systematically unaware of the full costs of government actions, including budgetary choices. The government in this case may strategically manipulate the fiscal institutions in order to strengthen its position, especially when it is trying to secure re-election. Our general framework is that of a monopoly held by the government, though one dependent on its having a minimum level of popular support. Such a monopolistic government will deviate from the median position, i.e., from the outcome that arises under perfect competition in a two-party system. It's rent consists of the hoped-for achievement of its ideological goals. The leeway for achieving such rent is restricted, however, as the government has to secure it re-election. One way to do this is to exploit fiscal illusions held by the voters.

We chose the federal government of Australia as our test case because the governing party, once elected, has a considerable amount of leeway to use fiscal policy instruments for its ideological purposes, which are very different for the two types of parties, and for exploiting fiscal misper-

ceptions when working towards its re-election.

Empirical evidence of systematic misperceptions of the cost of government is offered by the results of the government's popularity function. The findings indicate that voters are unaware of the full cost of government activities when indirect and more hidden revenue items are used for financing. There is also evidence that some spending items are favorably perceived, which accords with the findings of survey research. It remains open, however, to what degree this is due to preferences and to what degree it is due to misperceptions. This plus the fact that voters discount past government activities provide opportunities for the government to behave as if it had a monopoly position. As our empirical results show, significant differences between the two parties' use of fiscal instruments for ideological purposes do appear after an election when the winning party feels itself to be in a relatively secure position. However, when trying to secure re-election each government deviates from the pursuit of its ideological goals if necessary and uses fiscal instruments in a predictable fashion in order to exploit misperceptions. It is also shown that before an election a government will try to keep the fiscal burden as low as possible in the eyes of the voters and will undertake an expansionary spending policy in order to improve the general economic situation. After an election, changes are made in fiscal policy to create a less favorable perception on the voters' part so that the government can again have the opportunity to influence the voters' perception towards a more favorable view as the next election approaches.

Our starting point of a monopolistic government that uses its fiscal policy instruments in a strategic way seems to us to be important. Not only is the model's predictive ability quite good, we believe the whole approach to be worthy of further consideration. Monopoly government was, of course, already the framework for the debate concerning fiscal illusion in the classic continental European literature.[28] Such a framework, based on the idea of *dissimilarities* rather than similarities between politics and the perfect market, might serve to provide new and fruitful insights.

## Footnotes

* Both authors are at the University of Zürich, Switzerland. They would like to thank Domenico Da-Empoli, Giovanni Demaria, Bruno S Frey, Bernard Jurion, Gebhard Kirchgässner, Pierre Pestieau, Peter Zweifel and the participants of the fourth Arne Ryde Symposium on "Theories of Economic Institutions" in Lund 1979 for most helpful criticism and suggestion and Sandra Stuber for editing the English text.

1  Especially Stigler (1961), Telser (1966) and Nelson (1970).

2  See Nelson (1974), where the distinction is more fully elaborated.

3  This reasoning leads to a paradox in which voting appears not to be rational, but yet a considerable number of voters do vote. The problem we have set ourselves is to examine the rational ignorance argument and so the voting paradox is ignored here.

4  The general framework used here was developed by Frey and Lau (1968). Its usefulness for empirical research in the polito-economic context has been demonstrated by Frey and Schneider (1978a, 1978b) and Schneider (1978).

5  The Australian federal government is particularly suitable for our examination as it controls over 70 % of all public revenues and expenditures and also has the power to make rapid discretionary changes on the revenue side, i.e., changes in tax rate, tax deductions, tax base and tax rebates, and, to a lesser extent, in public expenditures. The Central Bank which in most Western democracies can restrict government policy does not play such an independent role here as it is part of the Ministry of Finance.

6  For a detailed discussion of the various cost arguments see Buchanan (1967, ch 10), Bartlett (1973, ch 8), and Pommerehne and Schneider (1978).

7  For the case of the United States see e.g. Katona (1975, pp 353 ff).

8  Even in the rare instances in which taxes are earmarked for specific public goods, such as highway construction and maintenance, the corresponding sales and excise taxes are often supplemented through a cross subsidization system which makes it almost impossible to derive even an accounting cost assignment of these expenditures to the individual.

9  Among the various possibilities for checking for voters' misperception are: (i) questionnaires on the awareness of the individuals marginal tax burden, which, however, are not taken on a continuous basis; (ii) analysis of actual-voters'-behavior at national elections, but as elections are held on only every 2 1/2 to 3 years, there are too few observations available to allow a quantitative analysis; (iii) analysis of surveys of the voters' evaluation of the government's economic performance and of their stated voting intentions. As there are monthly data available for these last two types of surveys going back to January 1970 we adopt this method.

10  Government popularity is measured by data series regularly collected by the Australian Gallup Poll showing the proportion of citizens "willing to vote for the Australian government at a federal election" at that point. These data and those for perceived government economic performance were given to us by Roger Douglas and Chris Goodrich to whom we would like to express our thanks.

11  All economic data and those for public revenues and expenditures were provided by Ernestine Gross and William S Hogan. We are especially grateful for their generous help, discussion and clarification of our questions on Australian institutional arrangements.

12  A three-month lag was chosen because the perception of a change in the economic situation is assumed to need about three months. However, when different lag structures, including weighted lags of up to one year, were used no major significant improvement occurred in the following simultaneously estimated equations.

13  As we used a unique constant term we had to drop one item share from the revenue and expenditure variables. In the case of the revenue side, we dropped the share of customs and duties and for the spending side, the share of defense expenditure. When these shares are included and others are dropped there is no significant change in our general results.

14  Similarly, Paldam and Schneider (1980) have shown for Denmark, and Pissarides (1980) has suggested for the United Kingdom, that, in explaining current government popularity, the variables that capture the voters' felt revenue burden should be considered with much more emphasis, in addition to the more classic macroeconomic variables.

15  See for instance Goetz (1977) and the literature quoted in footnote 6.

16  A similar result showing much bigger impact of the revenue side as compared to the spending side is also found for other countries, for example, for the United States see Niskanen (1979).

17  If the government is subject to a permanent, strongly binding re-election constraint, it must pursue a vote-maximizing policy in order to stay in power, i.e., it must behave in the same way as it would under a system of perfect party competition. In this case, as is well-known, government will use its various fiscal instruments to equalize the marginal vote gain of each individual instruments. Under simple majority voting and when there are only two parties this will lead to the median outcome.

18  As it is not possible in our framework to explicitly derive the ideological goals we have concentrated on the ideological preferences regarding fiscal policy goals and instruments as they are described in official party programs.

19  The method used here to measure the ideological preferences regarding fiscal instruments is rather simple. The constant term is broken into two dum-

my variables in order to capture the different uses of instruments by Country-Liberal and Labor governments with a shift parameter. We assume that the ceteris paribus conditions are fulfilled so that ideological differences can be measured by a broken intercept.

[20] When a longer or a more complex lag structure is applied the estimation and the forecast results do not improve significantly.

[21] If the two variables for securing re-election are split up for the two types of government both can be seen to undertake the same revenue policy. As the results do not differ significantly the variables for each revenue item are linked together.

[22] Additional empirical evidence for this argument is gained when we carry out a regression with the lagged endogenous variables of the various revenue and spending items (in billions of Australian dollars), thus capturing only the legal and administrative influences. The mean of the explained variance ($\bar{R}^2$) is then 82.5 % for the revenue items but is 93.1 % for expenditure by categories.

[23] For a more detailed analysis of this relationship in Australia over the longer period 1959 through 1978 see Schneider and Pommerehne (1980). Related studies done for other democracies have come to the same result; for a survey see Pommerehne, Schneider and Lafay (1981).

[24] We will not hypothesize as to what kinds of government expenditure will actually be increased most. The reason is that in some cases it may suffice to only announce additional spending programs in order to gain marginal votes. Moreover, there is a trade-off between the impact of such an announcement and the effects of the actual spending policy in stimulating the economy.

[25] An additional reason for such a policy is that the actual rate of inflation can be made to be quite different from that the voters expect in the hope that inflationary expectations can be reduced before the next election. Moreover, as Sjaastad (1976) and Johnson (1977) have shown, the government will gain more "inflation tax revenue" if it deliberately decieves the public as to its inflationary intentions and systematically varies the inflation rate than if it were to install and maintain a steady rate of inflation at a rate that would maximize tax proceeds under the conditions of a correctly expected steady-state rate of inflation.

[26] The various revenue and spending items are grouped mainly for data reasons as in equation 2.

[27] As the empirical findings for the influence of administrative and legal constraints are very similar to those discussed in Part II, they are not discussed here. Moreover, due to the lack of observations for the Labor government, we cannot interpret the ideological use of fiscal instruments. But it can be seen from the results that the ideological use of instruments by a Country-Liberal government also held over the longer period considered here.

[28] Perhaps the best example is Puviani's work (1896). For further references to classic work see Buchanan (1967, ch 10) and Pommerehne and Schneider (1978). This framework could also be one that stresses, among other things, the cartel-like characteristics of the behavior of political parties over a view emphasizing intense competition. While it would seem to be natural to also apply other economic models than that of competitive duopoly for the theory of political process, only a few attempts at such extensions have been made so far, see e.g. Spindler (1978).

## Bibliography

Barry, P F and C W Guille, (1976), "The Australian Business Cycle and Internal Cyclical Linkages 1959–1974", Economic Record, 52, pp 137–156.

Bartlett, R, (1973), Economic Foundations of Political Power, Free Press, New York.

Buchanan, J M, (1967), Public Finance in Democratic Process, University of North-Carolina Press, Chapel Hill.

Darby, M R and E Karni, (1973), "Free Competition and the Optimal Amount of Fraud", Journal of Law and Economics, 16, pp 67–88.

Downs, A, (1957), An Economic Theory of Democracy, Harper and Row, New York.

Freys, B S and L J Lau, (1968), "Towards a Mathematical Model of Government Behavior", Zeitschrift für Nationalökonomie, 28, pp 355–380.

Frey, B S and F Schneider, (1978a), "An Empirical Study of Politico-Eonomic Interactions in the U S", Review of Economics and Statistics, 60, pp 174–183.

Frey, B S and F Schneider, (1978b), "A Politico-Economic Model of the United Kingdom", Economic Journal, 88, pp 243–253.

Goetz, C J, (1977), "Fiscal Illusions in State and Local Finance", in: T E Borcherding, ed, Budgets and Bureaucrats: the Sources of Government Growth, Duke University Press, Durham, pp 176–187.

Johnson, H G, (1977), "A Note on the Dishonest Government and the Inflation Tax", Journal of Monetary Economics, 3, pp 375–377.

Katona, G, (1975), Psychological Economics, Elsevier, New York, Oxford, Amsterdam.

Nelson, P, (1970), "Information and Consumer Behavior", Journal of Political Economy, 78, pp 311–329.

Nelson, P, (1974), "Advertising as Information", Journal of Political Economy, 82, pp 729–754.

Niskanen, W A, (1979), "Economic and Fiscal Effects on the Popular Vote for the President", in: D W Rae and T J Eismeier, eds, Public Policy and Public Choice, Sage, Beverly Hills and London, pp 93–120.

Paldam, M and F Schneider, (1980), "The Macro-Economic Aspects of Government and Opposition Popularity in Denmark 1957–78", Nationaløkonomisk Tidskrift, 118, pp 149–170.

Pissarides, C A, (1980), "British Government Popularity and Economic Performance", *Economic Journal,* 90, pp 569–581.

Pommerehne, W W and F Schneider, (1978), "Fiscal Illusion, Political Institutions, and Local Public Spending, Kyklos, 31, pp 381–408.

Pommerehne, W W, F Schneider and J-D Lafay, (1981), "Les interactions entre economie et politique: synthès des analyses théoriques et empiriques", *Revue Economique,* 31, pp 110–162.

Puviani, A, (1896), *Teoria della illusione nelle entrate pubbliche,* Sandron, Milano, Napoli, Perugia.

Schneider, F, (1978), *Politisch-ökonomische Modelle,* Hain, Königstein.

Schneider, F and W W Pommerehne, (1980), "Politico-Economic Interactions in Australia: Some Empirical Evidence", *Economic Record,* 56, pp 113–131.

Sjaastad, L A, (1976), "Why Stable Inflation Fails: An Essay in Political Economy", in: M Parkin and L Zis, eds, *Inflation in the World Economy,* University of Manchester Press and University of Toronto Press, Manchester and Toronto, pp 73–86.

Spindler, Z A, (1978), "Oligopolistic Behavior and the Theory of Representative Democracy, Mimeo, Simon Frazer University.

Stigler, G, (1961), "The Economics of Information", *Journal of Political Economy,* 69, pp 213–225.

Telser, G, (1966), "Supply and Demand of Advertising Messages", *American Economic Review, Papers and Proceedings* 56, pp 457–466.

# Welfare Economics and the Welfare State
BY THOMAS WILSON*

*This paper discusses the role of the economist in making normative statements about the welfare state. The familiar difficulties of interpersonal comparisons and value judgement are examined. It is argued that the economist's role, though limited, is an important one not only in discussing the general economic implications of welfare expenditures but in analysing the scope that is permitted for the expression of individual preferences.*

## I

It would be only natural if those unfamiliar with economic literature were to suppose that there is an intimate association between "welfare economics" and "the welfare state". For such an association might be suggested by the repetition of the word "welfare", first in a technical and then in a colloquial sense. Indeed the welfare state might even be regarded as an institutional application of welfare theory! Some surprise might then be occasioned by their being told that this is by no means the case and that the links between welfare theory and the welfare services have been unsystematic and weak. Of course it is true that economists have not only embarked upon empirical investigations of the welfare state but have attempted to reach normative conclusions. The fact remains that the public provision of benefits, in cash and in kind, has been regarded by economists rather as a topic for specialised study that lies on the periphery of professional interest and has not been subjected to rigorous economic analysis as fully or as frequently as its quantitative importance would appear to warrant. At all events, this has been the position in Britain.

The explanation must be sought partly in the limited scope of welfare economics as it has evolved in modern times. "The welfare state" may be broadly defined as a group of official policies designed to provide certain categories of people with incomes, in cash and kind, in excess of the *current* contributions to production. Thus the study of the welfare state – unless so conducted as to be strictly empirical without normative conclusions – would seem to confront economists with precisely those problems of interpersonal comparisons of utility and of value judgements which

Paretian welfare economics has been designed to avoid. When cardinal utility was thought to be a respectable concept and its maximisation in a community was regarded as an objective that no-one could reasonably question, economists were prepared to commit themselves to recommendations about tax policies that were based on the distributional effects these policies could be expected to have. Admittedly not so much was said about expenditure – not, at least, in the Anglo-Saxon literature. But benefits to identifiable individuals – that is to say, benefits other than those derived from pure public goods – can be regarded as taxes with a change of algebraic sign. The earlier cardinalist literature could therefore be said to cover modern welfare benefits as well as taxes. When, however, this whole theoretical structure was blasted and flattened by the positivist wind that blew so strongly for so long a period, it became difficult to maintain that economists could make relevant recommendations without going beyond what appeared to be their legitimate domain. With interpersonal comparisons of utility ruled out and with moral value judgement avoided, the welfare analysis still deemed to be permissible had, naturally, much less to say that was relevant to the assessment of welfare policies of a redistributive nature. It is true that the restrictions thus imposed have not affected in the same way all the different aspects of policy with which economists are concerned. Thus the recommendations made about macro-economic policy do not seem to be much inhibited by the fact that interpersonal comparisons of utility and ethical value judgements are often embodied in them. For, though present, these difficult matters are decently concealed. It is a different matter when

distribution is a central issue, as in the case of the welfare state.

It is scarcely necessary to add that the difficulty of handling distributional issues is, indeed, so great that economic analysis can yield only provisional and qualified recommendations. Interpersonal comparisons of utility can, of course, be made; the difficulty is to apply some objective test and to achieve agreement. The basic value judgements embodied in distributional recommendations may also be conflicting. A conflict of basic value judgements may be said to occur when different recommendations would be made by different people even if they were in complete agreement about the facts of the situation. Of course it is the case that such difficulties limit the scope of welfare economics. The question that remains is whether these obstacles are so central and so immovable as to justify the extent to which the welfare services in general, and cash benefits in particular, have been neglected in welfare economics.

First, it may be asked whether the Paretian analysis itself, when amended to allow for interdependent utility functions, may be so extended as to permit relevant recommendations to be made. Secondly, it may be suggested that economists should be prepared to break out of the Paretian confines altogether, to commit themselves more boldly to expressions of opinion about justice or even to claim that cardinal utility is, after all, a meaningful and a useful concept.

Whether or not such explorations prove to be successful, it remains true that there is scope for the fuller and more persistent application of the economist's more traditional skills. After all, the basic questions about the allocation of resources need to be asked in this context as in others. How is the allocation to the welfare services as a whole determined? What determines the allocation between different parts of the total welfare budget? Is the use of resources reasonably efficient, given the preference patterns? Are people in a position to express rational preferences between different alternatives? What is the justification for the compulsion that underlies all these schemes and for the uniformity of provision which they often impose? And so on. There is no need to add to the list at this stage. Nor is there any need to say that such far-reaching questions cannot be satisfactorily pursued within the scope of a single paper. It must suffice to develop some of the more general points a little further in the next part, in the following one to provide some illustrations,

and then, finally, to put forward some conclusions.

## II

Analytical welfare economics is firmly individualistic. It is so, first of all, in the anti-metaphysical sense that social welfare is identified with the welfare of the individuals who comprise society, not with that of some mystical entity such as the Nation, the People or the Proletariat. Individualism carried only to this extent would impose no obstacle to the analysis of the welfare state; but it may be pushed much further. Thus, in the standard version of Paretian welfare economics, utility functions are independent so that each person "exists in an isolated cell connected with the rest of the world only through the exchange of goods and services." See Vickery (1973, p 37). At first glance this would not appear to provide a promising basis for the study of the welfare state, but it is necessary to reflect upon the reasons why an assumption of "isolated cells", may be made. First of all, it should be noted that the reference to the exchange of goods and services is crucial. With an elaborate division of labour, everyone has a particular area of responsibility as producer or consumer which must be his or her own particular concern. Special circumstances apart, you are under no obligation to accept a bad bargain from me as though you were assuming my responsibilities as well as your own. On the contrary, you should try to get as good a bargain as you can and you can feel that, in doing so, you will usually be enhancing social welfare. For this is how information about preferences and costs is conveyed; this is how efficiency is rewarded and inefficiency penalised. (We may note in passing that much the same can be said about the Socialist economies as well as the Western ones.) In short we must distinguish between non-tuism, C. C. Wickstead (1833) called it, and selfishness. One may be behaving in a non-tuistic way but one's objectives may be selfish or altruistic or a combination of both. See MacKean (1975) and Wilson (1976). It was entirely proper to base welfare economics on the assumption of independent utility functions in the sense of non-tuism in so far as this welfare analysis was concerned with production and exchange. Public finance, including the welfare services, raises other issues.

Secondly, the assumption that people occupy isolated cells may be based on the quite different Hobbesian assumption of egoism over the whole

range of personal activity. This is a much stronger assumption but is by no means inherently inconsistent with support for any form of welfare state. In giving such support people might only be resorting to what they regarded as one of the available methods for advancing their self-interest. In doing so they would, it is true, be endorsing the use of compulsion and accepting some measure of uniformity in the provision of services which might conflict with their own preferences; but they might feel that, on balance, the package was worthwhile. It might be so regarded even by those who expected that their income class would be made to pay more than its proportionate share because, as individuals, they would have protection against risks. Those who believed – rightly or wrongly – that other classes would have to meet proportionately more of the cost would have a still stronger egotistical reason for giving their support. Given this starting point in egoism, the economist would have the role of assessing how far the pursuit of self-interest in the case of particular welfare services was really enlightened; for people might, of course, be mistaken about the means they had chosen to advance their objectives.

Although, for the purposes of an exercise of this kind, economists might accept the assumption of completely egotistical behaviour, they would be going beyond their professional competence if they themselves were to assert that such an assumption is a fully realistic statement of human motivation. For it could be objected, whether on the basis of psychological enquiry or simply of ordinary experience, that people do not live in isolated cells, that interpersonal comparisons of utility can be made and are, in fact, made and, therefore, that utility functions are not independent. To recognise that this is so does not preclude acceptance of the Paretian proposition that situation A is better than situation B if individual i is better off and no one else is worse off, but "better off" and "worse off" must be taken to refer not only to supplies of goods and services but, in the most comprehensive sense, to all conceivable considerations. The latter must include any feelings of guilt on the part of the more prosperous – the "middle class conscience" – and feelings of resentment on the part of those who believe themselves to be relatively deprived.

Interdependence in utility functions provides an adequate explanation of altruism, a point that was made more elegantly by Adam Smith himself. It has been maintained, however, that interde-

pendence also provides an explanation of compulsory fiscal transfers. See Hochman and Rogers (1969). There are various reasons why compulsion might be introduced. Thus, it might be held that in this way the total sum to be transferred could be better assessed and determined beforehand and the burden more fairly distributed. We could therefore envisage donors expressing their preferences by agreeing voluntarily to a social contract that would provide for compulsory transfers. The practical difficulties would, of course, be immense with voting in the form of referenda on specific issues and with the franchise confined to potential donors. Moreover, if the contract were to be genuinely voluntary so that it could be defended to a Paretian welfare economist as an expression of what were basically unforced preferences, no-one should be obliged to give more than he wished to give. An important conclusion follows. It is that even if we were to relax Wicksell's unanimity rule a little, as he himself was prepared to do, it is scarcely conceivable that transfers would take place on the large scale to which we have become accustomed. It should be observed that, in this situation, those who were potential donors would not be making their decisions behind a Rawlsian "veil of ignorance" and might feel, in many cases, that there was not much likelihood that they themselves would have to rely significantly upon the receipt of transfers. Though moved to some extent by sympathy, benevolence or a sense of guilt, they could also be expected to accord heavy weighting to their own self-interest or to the welfare of those who were close dependents.

The method of approach we have just described is in any case open to the objection that it is too conservative in that excessive stress is placed on voluntary modifications of an initial income distribution which may have been a far from satisfactory one. Adam Smith's "Impartial Spectator" would surely have expected more than this! It can be pointed out that this particular distribution is, in principle, only one of many possible distributions. If movement from it is only to be made to the extent that this is sanctioned by those who will voluntarily surrender some goods and services, then the whole welfare system is still only partially ordered. The *optimum optimorum* has not yet been reached. In reply it may be said that one must always start from some historical situation and that, in any case, it is foolish to talk about an *optimum optimorum* which will not be attained and could not even be defined without embarking upon an ethical investigation into so-

cial justice. It does not follow, however, that the initial distribution is in some sense superior to all others.

Some economists, in their impatience with Paretian inhibitions, appear to have been quite prepared to direct their attention to questions of social justice. The interest aroused by Rawls' *Theory of Justice* (1971) was evidence of this dissatisfaction and readiness to consider new approaches. Unfortunately the difficulties have remained. Rawls' views were widely welcomed; they were also widely criticised. This outcome was scarcely surprising. This is not a field of speculation where agreement is likely to be reached. An economist who decides nevertheless to explore this field must be fully explicit about what he is doing and careful not to represent as conclusions of economic science, recommendations that are substantially derived from non-economic considerations.

## III

The welfare state, as defined above, includes a wide range of measures. This range is sometimes limited by convention. For example, the EEC includes cash benefits and health benefits in kind; but educational benefits in cash and kind could also be included and so could housing subsidies in so far as these are not covered by the cash benefits. Rather than attempt to comment on so wide a range of provisions within the compass of a short paper, it may be more illuminating to take an example and the one I shall choose will be the provision of benefits to the elderly which account for much the larger part of cash benefits in nearly all countries.

Why should such assistance be thought to be necessary? Why should people in rich developed countries not be left to make provision for themselves? The number likely to be destitute could be expected to be quite substantially less than the present number of pensioners if people were taxed less and left to look after themselves. But some would be in trouble as a consequence of invalidity or ill-luck in the management of their affairs or simply from improvidence. We can reasonably assert that there would be almost universal support for the view that no elderly person should be left totally destitute partly because utility functions are interdependent and partly because self-interest may lead people to endorse the provision of benefits which, in the unforeseeable future, they might need. It is true that the families of the needy could be expected to come to the rescue as do

the extended families in less developed countries, but family responsibility has been undermined in developed countries – though it is fair to add that this is a consequence as well as a cause of assistance from the state.

The assistance thus deemed to be necessary could be provided only subject to means-test. Total expenditure would then be substantially less than it is under an official pension scheme because assistance would not be provided to those who did not need it – apart from errors in administration. The possibility of relying upon means-tested assistance may seem rather academic in Europe and North America but it should be noted that this has been a real issue in Australia. Does welfare economics help in making a choice? There are the obvious points that means-tests weaken the incentive to help oneself by working and saving. On these points at least the economist should be able to give clear advice of a quantitative nature but, in fact, finds it hard to do so. Empirical matters of this kind which are within our competence in principle are not always so in practice! Then there is a more basic philosophical point. Means-tests may undermine self-respect with the result that some people will be tempted to be free-riders and others will feel a sense of stigma.

Economists must be careful at this point. Welfare economics is still basically utilitarian even although cardinalism has been abandoned and "the greatest welfare of the greatest number" is seen to be not only vague but to embody a value judgement. It should not be forgotten, however, that in assessing any public policy, account can also be taken of its effect on personality, on the sense of duty and the sense of obligation. It is not, of course, the job of the economist to decide what weighting should be given to deontological considerations of this kind but he must be careful not to adopt, implicitly, a particular philosophical position by leaving them out of account.

If it were, indeed, possible to devise a social insurance scheme under which everyone paid for what he got, then the distributional issues which cause so much difficulty in welfare economics would appear, at least at first glance, to be bypassed. This, of course, is what Beveridge (1942) recommended. There should be flat-rate benefits for flat-rate contributions with the need for means-tested assistance – "public charity" – reduced to a small number of difficult cases. It is true that each person's benefits would not correspond precisely to his contributions for this was to be an insurance

scheme. Moreover, it was to be a national scheme which did not allow, as private insurance would do, for the fact that certain categories of people were more subject to risk than others. Some distributional changes would take place but, given all the assumptions of the proposal, their quantitative importance would be limited and also justified if the basis of the scheme were generally accepted. Beveridge, however, was in various ways inconsistent in applying this recommendation that people should pay for what they got. Thus, he proposed that there should be benefits for the dependents of the insured and this implied transfers. In practice, nearly all schemes do, in fact, provide for dependents although the German pension was for some time an exception. This issue has been taken up recently in the USA, notably by Derthick (1979) who maintains that dependents should be excluded from social insurance and provided for by means-tested assistance. Does welfare economics help at this point? Only, perhaps, to the extent of endorsing the general case for the consistent application of concepts. That is to say, if we are talking about social insurance we should not smuggle in something else. But "social insurance" may now be so muddied a concept for quite a number of reasons that it is pointless to try to purify it.

If it is assumed that some national minimum is to be provided, as of right or subject to means-tests, it is still necessary to determine the level at which it should be set. There is no definite scientific way of making an assessment and judgement is always involved. What can be said with some confidence is that the higher the level that is being considered, the greater the likelihood of disagreement. The Hochman-Rogers (1969) approach would no doubt yield nearly unanimous support for protection against the desperate extremities of total destitution; but views will differ more, compulsion will matter more and political weight will become more important when the minimum is far above what would be needed as protection against destitution and provides for at least some conventional luxuries. There is the further question as to how often the basic minimum, once fixed, should be altered. Interdependence in utility standards helps to explain the fact that "poverty" is a relative concept. But how relative? How often should the standard be changed? Should it rise with rising real wages as was the case until recently in Britain – which is one of the reasons why neither growth nor the welfare state has seemed to be successful in abolishing

poverty! The economist can give no definite answers to these questions but he, for this part, can ask some empirical questions. How has the minimum been determined at a particular point in time? By whose authority? With what attempt to test public preferences? With what allowance for wider consequences in the economy? He may also perform the more humble but exceedingly important role of combating the misunderstanding caused by ambiguous statements about "poverty".

Flat-rate benefits are now the exception for pensioners and, in the countries where such benefits still survive – as in Sweden and Britain – there are also supplementary pensions which are graduated. See Wilson (1974). Graduation clearly raises important issues. What is the justification for official arrangements, backed by compulsion, for carrying over into retirement some of the inequality of working life? It is true that this transmission of inequality takes place only within limits, for the benefits normally have maxima and minima and, furthermore, a graduated scheme may be kinked in such a way as to favour the poorer pensioners – as in the USA. But it is still necessary to ask why there should be graduation at all. If one is prepared to adopt a bold utilitarian attitude, the beginnings of an answer can be given. For it can be held that what should concern us is the distribution of welfare, not the means to welfare. That is to say, allowance should be made for the fact that the capacity for enjoyment differs from person to person. For example, Sen (1973, p 87) has argued that an invalid can be expected to derive less utility from a given income than a normal healthy person and should, in fairness, be given more for this reason. When the application of this argument is extended it becomes apparent that it has strong conservative implications, for the satisfaction a person can obtain from a given income will depend partly upon the standard of living to which he has been accustomed. Habit is far wider in its relevance than the special case of invalidity to which Sen refers.

There is, however, no real need for economists to become quite so deeply involved in interpersonal comparisons of utility. Such comparisons can, of course, be made but are hard to check. As has been suggested, there appears to be a sufficient consensus to warrant support for providing some minimum standard, some floor below which no one should fall; but we should be plunged into controversy and baffled by uncertainty about psy-

chological states if we were to attempt to devise a distribution of income that allowed for differences in the capacity to enjoy. Of course, it is to be expected that people with higher incomes at work will want to have higher incomes in retirement, but it is up to them to make provision accordingly. Compulsory contributions for official benefits impose a degree of uniformity in life-cycle provision which is not easily justified in principle.

There is an analogy here with the debate about the case for providing benefits in cash or in kind. Welfare theory would seem to support the former because preferences differ and satisfaction will be greater if this is recognised – the utilitarian point. To do otherwise is to subject people to dubious interpersonal comparisons of utility made by some outside authority on their behalf. Moreover the exercise of freedom of choice may also be preferred on deontological grounds. It is true that a strong case can be made for a more paternalistic and a more egalitarian approach in the case of health services or, at all events, of that part of the health service which is crucial to the preservation of life or the mitigation of serious suffering, but it becomes a lot less plausible in less desperate cases. The extension of this line of reasoning to graduated pensions fails, however, to carry the same conviction.

The benefits provided in some countries to the elderly are now so large as to impose a heavy burden on the working population which is beginning to be resented. This would appear to be the case in Sweden. See Wilson (1979). The aim in some countries is to provide about three-quarters in income from work of the average wage-earner which, when allowance is made for lower costs after retirement and for tax, would permit something like the standard of living of working life to be maintained from state and occupational pensions apart from any dissaving of previous private accumulations. Those with less than average earnings may even be better off in retirement than they were when working and perhaps supporting a family. It is true that we have not moved so far along this path in Britain and it is not planned to do so; but a number of continental countries have gone a long way, or will go a long way when existing schemes are fully mature. Moreover, the elderly make a heavy claim on health services. It is quite reasonable to ask whether those still at work, especially those with children to support, are not being expected to do too much.

It may be objected that the economist, whatever views he may hold as a citizen, has no basis

for appraising this scale of provision for the elderly, either favourably or unfavourably, so long as he confines himself to his own subject. If "society" has chosen such a policy, why should he object? In fact, however, he has still something of importance to say. First, he can point out that large unfunded schemes may have a bad effect on saving with damaging consequences for growth. It does not follow that a return to funding should be recommended for this would not be realistic. No working population can be expected to provide *both* for those who are currently beneficiaries on a pay-as-you-go basis (because there is no fund on which to call) and also to contribute to a fund for their own future. Funding for official schemes is now relevant only when new schemes are introduced. Even if this were not so and all benefits could be funded, the size of the fund required would be enormous and the rate of return on capital would then be low – a fact which, in itself, shows that beneficiaries may often be getting more than they paid for in the sense that the implicit return on their contributions is above the probable market rate that could be expected on a fund. The managers of so large a fund would also be in a position to exercise immense control over industry. Even the funded private schemes convey a great deal of power to the financial institutions which control them. This is now an issue in Britain. Is it really desirable to reduce so drastically the role of the private investor?

Secondly, the economist can ask whether the citizens of a country have been offered a clear and fair choice between the forced saving implied by both public and private schemes on the one hand and voluntary personal saving on the other – above what would be needed for some reasonable minimum. The fact that *employers* make large contributions confuses the issue. For people must often suppose that they are getting benefits at the expense of profits. But the sums involved are so large that the levy, whether for official or private schemes, must be passed on almost entirely as higher prices. Thus, the burden falls on the population at large as an indirect tax of uncertain incidence with no regard to equity. Moreover in so far as profits are sometimes squeezed, investment and output may suffer. Finally, the tax on labour warps the choice between factors. There can be no doubt that the employers' payroll tax is a bad tax and ought, if this were politically possible, to be scrapped. Hypothecated taxes are supposed to convey more clearly the sense that there is a budget constraint but this tax does

not do so and is inferior for this reason, and for the other reasons just mentioned, to finance from general revenue.

The fact remains that it would be foolish for any individual acting in isolation to opt out of these schemes, if the choice were given, unless compensated to the extent of all the payments that would otherwise be made on his behalf including the tax concessions. The only choice ever offered, as far as I am aware, is that between public and private schemes, and this is not enough. It cannot be said, therefore, that the present system reflects preferences freely expressed between fairly presented alternatives. We have spent much time discussing the imperfections of the market, but must not neglect the imperfections of the political process.

There is, of course, another reason why, in most countries, it would be foolish for any individual to opt out of an official pension scheme. This is the fact the such schemes are usually indexed and provide a measure of security against inflation that could not be obtained from holding a financial asset. To test preferences by offering the opportunity to opt out would be a pointless exercise unless some inflation-proofed asset were provided as an alternative.

It may seem somewhat academic and fanciful even to consider the possibility of allowing people to opt out of all but the basic benefits. Yet in practical terms, this could be done subject to the condition that there could be no opting out and in and out again with changes of mind! The other reason for reflecting on this possibility is that, by doing so, one perceives quite clearly how warped the welfare state has become. There has been no question of allowing people to express their preferences as between different fairly balanced and fairly presented alternatives. In traditional welfare economics great stress is placed on the need to respect individual preferences and to bring them into the "right" relationship to costs; but in public provision, preferences receive too little attention. Graduated pensions are, of course, only one example.

## IV

Assar Lindbeck (1975) has observed that a large part of welfare expenditure in Sweden involves no vertical redistribution. This is also true of other countries, and the time has surely come to consider again the purposes that these vast programmes are meant to serve. Protection against "po-

verty" was once the central objective but it is now only one of several. When account is taken of the whole range of services provided by the welfare state, both in cash and in kind, it becomes obvious enough that large sums are paid over for benefits that the recipients are, in effect, paying for but have not freely chosen. Moreover, the paternalism which has accompanied and partly inspired the empire-building of politicians and bureaucrats can be regarded as unattractive from a deontological as well as a utilitarian point of view.

The stagflation with which the nineteen eighties have begun, together with ominous demographic movements in some countries, has led to growing concern about the current scale and possible expansion of welfare expenditure. In some countries (e.g., in Britain), certain changes have already been made that would have been considered politically impossible a few years ago. In particular, the rules for automatic indexation have been changed for some benefits or abandoned for others which means that the scope for dicretionary action in the future has been widened, notably discretionary action with regard to replacement ratios where this term means the ratio between benefits and previous income. Of course, there can never be any question of scrapping existing welfare structures and substituting something completely new. It is never possible to start building again on a cleared site. The fact remains that there is now some room for manoeuvre and the general economic situation requires that possibilities should not be neglected. The task before each country is therefore to assess the scope for change and to decide where change is practicable and could most usefully be made.

## Footnotes

\* Adam Smith Professor of Political Economy. University of Glasgow.

## References

Baumol, William J, (1975), "Business Responsibility and Economic Behaviour" in *Altruism, Morality and Economic Theory* edited by E S Phelps.

Beveridge, W H, (1942), *Social Insurance and Allied Services,* HMSO, Cmd, 6404.

Brittain, J A, (1972), *The Payroll Tax for Society Security,* Brookings Institution.

Buchanan, J, (1968), "Social Insurance in a Growing Economy: A Proposal for Radical Reform", *National Tax Journal.*

Derthick, M, (1979), *Policymaking for Social Security,* Brookings Institution, Washington, D C.

Hochman, H and Roger, J, (1969), "Pareto Optimal Distribution", *American Economic Review*.

Knoeber, C R, (1977), "The Demand for Government Transfers: A Look at the Recipients", *Public Choice*.

Lindbeck, A, (1975), *Inequality and Redistribution Policy Issues*, Seminar Paper no 44, Institute for International Economic Studies, Stockholm.

McKean, R N, (1975), "Economics of Trust, Altruism and Corporate Charity" in *Altruism, Morality and Economic Theory*, ed E S Phelps.

Heclo, H, (1974), *Modern Social Politics in Britain and Sweden*, Yale U P.

Rawls, J, (1974), *A Theory of Justice*, Harvard U P.

Sen, A, (1973), *On Economic Inequality*, Oxford U P.

Vickery, W S, (1973), "An Exchange of Questions between Economics and Philosophy", in *Economic Justice*, ed E S Phelps, Penguin Books, London.

Wickstead, P, (1833), *The Commonsense of Political Economy*.

Wilson, T (ed), (1974), *Pensions, Inflation and Growth*, Heinemann, London.

Wilson T, (1976), "Sympathy and Self-Interest", in *The Market and the State*, ed by T Wilson and A S Skinner, Oxford University Press.

Wilson, D J, (1979), *The Welfare State in Sweden*, Heinemann, London.
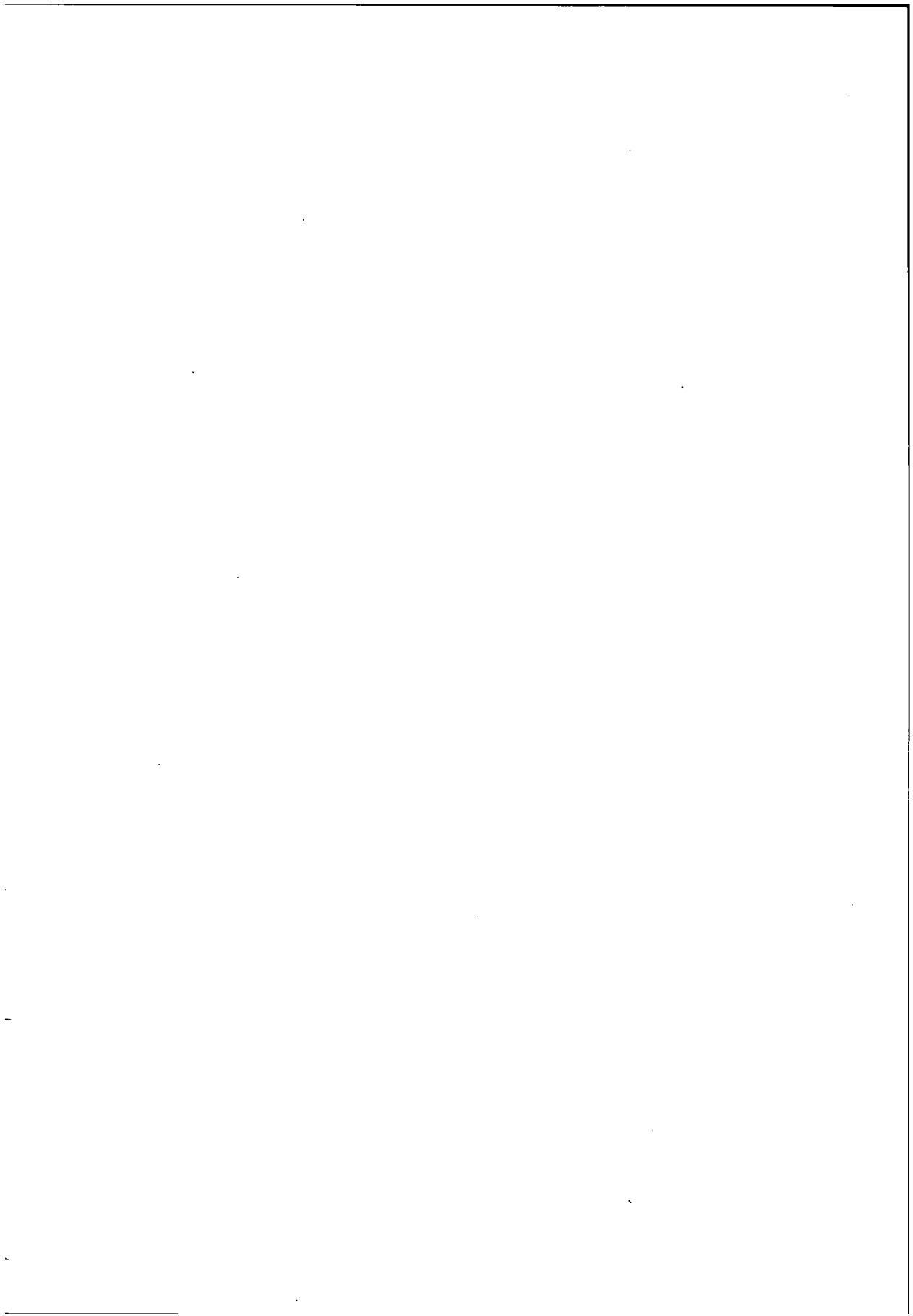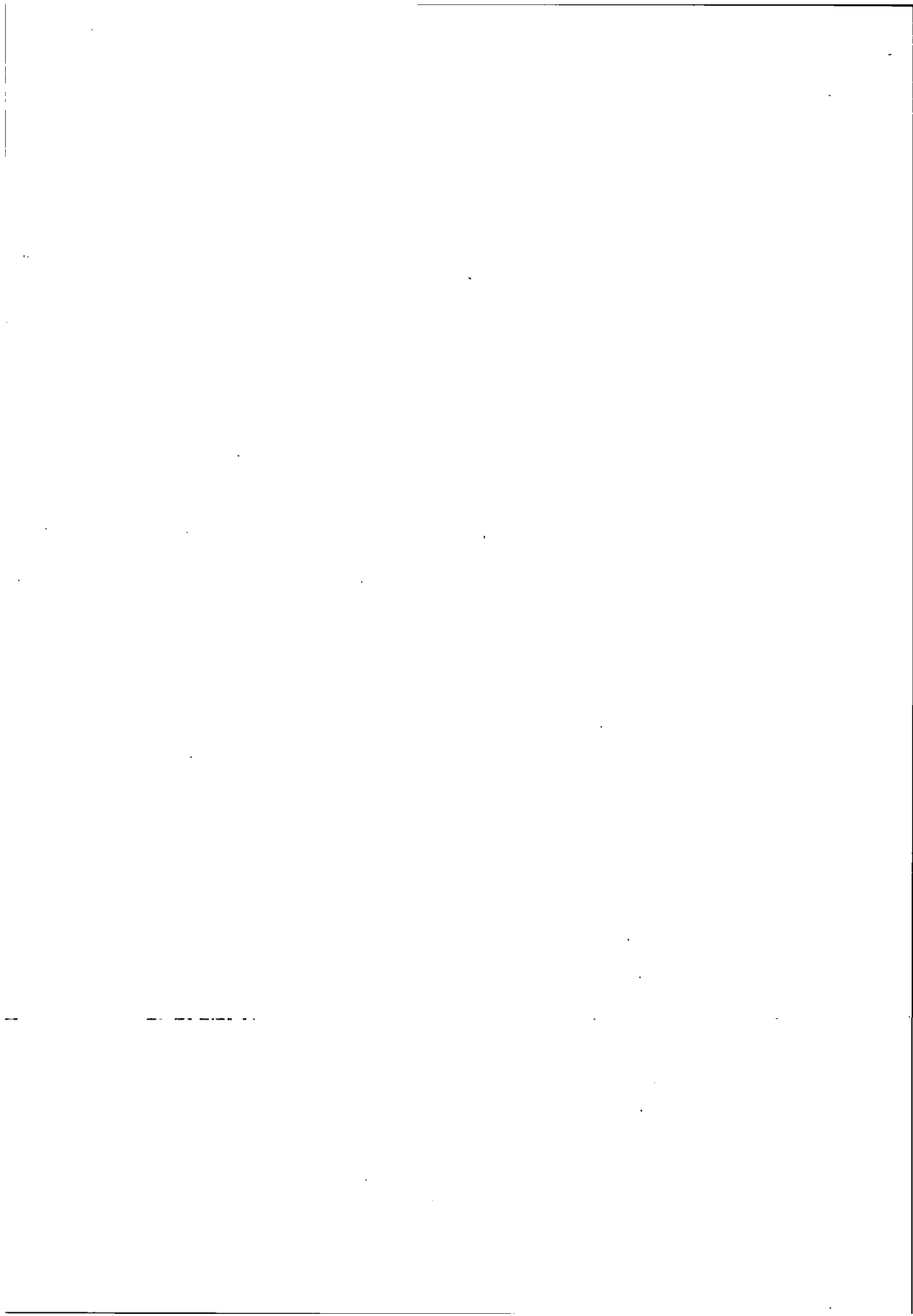
Copies of this special issue of the *Statsvetenskaplig Tidskrift* (The Swedish Journal of Political Science) can be ordered from Department of Economics, University of Lund, Box 5137, S-220 05 LUND, Sweden.

Price: 40 Swedish Crowns a copy.